

THE SCIENCE OF CLIMATE CHANGE

THE SCIENCE OF CLIMATE CHANGE

THE SCIENCE OF CLIMATE CHANGE

nature

THE COMING CLIMATE CRUNCH

- The trillionth
tonne of carbon
- How disastrous
can it get?
- Engineering
alternatives

NATURELIVE
Interviews



Abstractions



FIRST AUTHOR

Optical imaging is an indispensable tool for studying dynamic processes at high speeds and on micrometre scales. Unfortunately, conventional electronic cameras relying on charge-coupled devices (CCDs) cannot capture fast processes such as the dynamics of cell signalling with high sensitivity. On page 1145, Keisuke Goda at the University of California, Los Angeles, and his colleagues reveal a blueprint for the world's fastest video camera. Goda tells *Nature* that he is eagerly awaiting the scientific breakthroughs that this camera could bring.

Was your work motivated by a potential application or a technical challenge?

My adviser Bahram Jalali's group holds world records in speed for the analogue-to-digital conversion of electrical data and for optical spectroscopy — both using a technique that maps the optical spectrum into a data stream in real time and simultaneously amplifies it optically. We developed a new type of optical-imaging system called serial time-encoded amplified microscopy, or STEAM, which captures two-dimensional images at a frame speed of 163 nanoseconds and a shutter speed of 440 picoseconds — 1,000 times faster than CCDs. As a result, it has a broad range of applications, from imaging shock waves to microfluidics to chemical dynamics in living cells.

How did you get the idea for STEAM?

I went to the 2008 Optical Society of America conference in San Diego, California, where I got a hint that we might be able to achieve optical-image amplification by combining spectrally encoded imaging with our ultrafast spectroscopy approach. With the integrated technique, each pixel is encoded in a different frequency of the optical spectrum, which allows us to amplify the image optically. As a result, we can achieve imaging at high speeds without sacrificing sensitivity — which is what happens when you try to amplify an image electronically.

Are researchers clamouring to get this technology?

No one has to beat down my door to get it. Most components of STEAM are commercially available so anyone interested can build a STEAM camera themselves using our paper as a blueprint.

What is the next record to break?

There is always a new record to be broken, but this paper is not the end of this research. We want to keep increasing the frame rates and shutter speeds to make our camera better. I'm also pushing towards developing three-dimensional imaging at very high frame rates and shutter speeds. ■

MAKING THE PAPER

Luke Harmon

Evolution can put the squeeze back on the ecosystem.

Changes in ecological settings are thought to drive the evolution of one species into many — a process called adaptive radiation. Darwin's finches are a classic example, evolving into species with different beaks adapted to food availability in their island environments. However, few have considered the converse: that evolution might affect ecology. In an experimental demonstration of this idea, Luke Harmon of the University of Idaho in Moscow and his colleagues show that fish speciation can have profound effects on freshwater ecosystems.

Harmon credits his co-author Dolph Schluter of the University of British Columbia in Vancouver, Canada, with the idea of testing the reverse process. For the work, Harmon chose freshwater threespine sticklebacks (*Gasterosteus aculeatus*), which evolved from marine sticklebacks fairly recently — in the 10,000 years since the glaciers retreated from British Columbia. In some lakes, the sticklebacks evolved further into two separate subspecies: a bottom-dwelling fish that feeds on invertebrates in the mud, and an open-water fish that skims near the surface eating zooplankton. For Harmon, this represented the very first step of adaptive radiation — the divergence of one species into two. But what happens to a lake ecosystem as a result of that speciation?

Harmon's group created artificial ponds in 1,136-litre tanks and added fresh pond muck to deliver invertebrates and zooplankton. They then introduced either sticklebacks from single-species lakes, which Harmon calls "generalists", or the more specialized sticklebacks from lakes with two subspecies.

As the 10-week experiment progressed, some tanks bloomed with algae, whereas others showed limited growth. The researchers measured changes in plant and animal populations



in the tanks, but the results, Harmon says, were "pretty chaotic. We could tell that there was an effect, but we couldn't explain it."

One part of Harmon wanted to be done with it. "I could've written a paper at that point that said, 'Evolution matters, but in unpredictable ways,'" he says. The other part of him believed that important answers lay in his tanks' muddy waters, so he sought advice from Blake Matthews at the Swiss Federal Institute of Aquatic Science and Technology in Kastanienbaum, an expert in aquatic ecosystems. He pointed out dozens of physical measurements that Harmon's team had overlooked, such as the amount of dissolved organic carbon (DOC) and light transmission.

When the researchers ran the experiment a second time and included these measurements, they found that DOC was a key factor. DOC is essentially organic detritus and gives ponds their tea-like colour. In tanks with the one generalist fish species, most DOC molecules were small enough that light could penetrate to the bottom of the tank, and algae bloomed. In tanks with the two specialist subspecies, DOC molecules were larger, blocking out light. The authors suggest that differences in feeding between the generalists and the two specialists is what alters DOC size (see page 1167).

Harmon says he didn't appreciate how much "organisms can alter the physical structure of their environment through behaviours such as feeding". And he's not the only one. Lake water chemistry, he adds, "is one of those huge areas of ecology that most evolutionary biologists don't think about". ■

FROM THE BLOGOSPHERE

When did your career turn a corner? Was it a mentor's nudge, a flash of brilliance or an experimental second thought that became a lifelong pursuit? In the first instalment of the 'Turning Points' series in *Nature Cell Biology*, we learn how Gottfried Schatz was drawn into a career studying mitochondria by an unusual response to a postcard.

The Nautilus blog introduces

these "short autobiographical essays by leading scientists", which "feature a first-hand recounting of a pivotal event that shaped his or her scientific future" (<http://tinyurl.com/cjfmzt>). The journal's editors hope that some cell-biology folklore, normally shared only at the bar among a privileged few conference-goers, will be appreciated by larger numbers of early-career

scientists each month.

Schatz, emeritus biochemist at the University of Basel in Switzerland, tells how his career was kindled by a senior scientist's generosity (<http://tinyurl.com/coqlvy>).

If you would like to see a particular cell, molecular or developmental biologist featured in this series, please send your suggestion to cellbio@nature.com. ■

Visit Nautilus for regular news relevant to *Nature* authors ▶ <http://blogs.nature.com/nautilus> and see Peer-to-Peer for news for peer reviewers and about peer review ▶ <http://blogs.nature.com/peer-to-peer>.



Time to act

Without a solid commitment from the world's leaders, innovative ways to combat climate change are likely to come to nothing.

It is not too late yet — but we may be very close. The 500 billion tonnes of carbon that humans have added to the atmosphere lie heavily on the world, and the burden swells by at least 9 billion tonnes a year (see page 1117). If present trends continue, humankind will have emitted a trillion tonnes of carbon into the atmosphere well before 2050, and that could be enough to push the planet into the danger zone. And there is no reason to think that the pressure will stop then. The coal seams and tar sands of the world hold enough carbon for humankind to emit another trillion tonnes — and the apocalyptic scenarios extend from there (see page 1104).

Nations urgently need to cut their output of carbon dioxide. The difficulty of that task is manifest: emissions have continued to rise despite almost two decades of rhetoric, diplomacy and action on the matter. But that unhappy fact should not be taken as a licence for fatalism. Governments have a wide range of pollution-cutting tools at their command, most notably tradable permit regimes, taxes on fuels, regulations on power generation and energy efficiency, and subsidies for renewable energy and improved technologies. These tools can work if applied seriously — so citizens around the world must demand that seriousness from their leaders, both within their individual nations and in the international framework that will be discussed at the United Nations Climate Change Conference in Copenhagen this December.

As essential as it is, however, simply agreeing to cut emissions will not be enough. The fossil fuels burned up so far have already committed the world to a serious amount

"Even a complete halt to carbon pollution would not bring the world's temperatures down substantially for several centuries."

of climate change, even if carbon emissions were somehow to cease overnight (see page 1091). And given the current economic turmoil, the wherewithal to adapt to these changes is in short supply, especially among the world's poor nations. Adaptation measures will be needed in rich and poor countries alike — but those that have grown wealthy through the past emission of carbon have a moral duty to help those now threatened by that legacy (see page 1102).

The latest scientific research suggests that even a complete halt to carbon pollution would not bring the world's temperatures down substantially for several centuries. If further research reveals that a prolonged period of elevated temperatures would endanger the polar ice sheets, or otherwise destabilize the Earth system, nations may have to contemplate actively removing CO₂ from the atmosphere. Indeed, the United Nations Intergovernmental Panel on Climate Change is already developing scenarios for the idea that long-term safety may require sucking up carbon, and various innovators and entrepreneurs are developing technologies that might be able to accomplish that feat (see page 1094). At the moment, those technologies seem ruinously expensive and technically difficult. But if the very steep learning curve can be climbed, then the benefits will be great.

More radical still is the possibility of cooling the planet through some kind of 'geo-engineering' that would dim the incoming sunlight (see page 1097). The effects of such approaches are much more worrying than those of capturing carbon from the air, however. The cooling from geoengineering would not exactly balance the warming from greenhouse gases, which would cause complications even if the technology itself was feasible — something for which the evidence has been circumstantial, at best.

But discussions about the possibilities offered by geoengineering could also lull the world's leaders into complacency — if they lead them to believe that the technology

EDITORIAL

1077 **Time to act**

NEWS FEATURES

1091 **A burden beyond bearing**
Richard Monastersky

1094 **Sucking it up**
Nicola Jones

1097 **Great white hope**
Oliver Morton

COMMENTARY

1102 **Overshoot, adapt and recover**
Martin Parry, Jason Lowe & Clair Hanson

ESSAY

1104 **The worst-case scenario**
Stephen Schneider

BOOKS & ARTS

1107 **Could climate change capitalism?**
Robert Costanza

1112 **Managing nature as Earth warms**
Jessica J. Hellmann

NEWS & VIEWS

1117 **Too much of a bad thing**
Gavin Schmidt & David Archer

LETTERS

1158 **Greenhouse-gas emission targets for limiting global warming to 2 °C**
Malte Meinshausen *et al.*

1163 **Warming caused by cumulative carbon emissions towards the trillionth tonne**
Myles R. Allen *et al.*



For podcast and more online extras see www.nature.com/climatecrunch

will provide an escape hatch if the climate ever does reach a tipping point. This does not mean that the discussions should be avoided, but rather that the speculations need to be backed up with a solid body of research. Moreover, geoengineering research should be framed not as a hope for *deus ex machina* fixes to sudden global deterioration, but as a palliative cushion for the worst excesses of the peak years that are inevitable even after emissions start to be cut. A world slightly shaded from the Sun while its carbon levels are brought down by means of active capture would be a strangely unnatural place — but not necessarily a bad one, compared with the alternatives.

Research on local and regional interventions to cool Earth should be undertaken now — ideally in a way that provides basic information to climate researchers. Most sciences are in the habit of poking that which they study to gauge its response. Climate researchers lack such a tradition, and might have something to gain by starting

one up. The attention to far-off goals, however, must not obscure short-term opportunities. In addition to cutting CO₂ emissions, global leaders should curb the release of other substances warming the climate, notably methane and soot, also known as 'black carbon'. Tackling such pollutants will bring other benefits, too, such as reducing the respiratory problems associated with cooking over smoky fires and with high levels of tropospheric ozone.

With so many challenges still to be faced, the climate problem may seem insurmountable. But there is still time left to act, and there is hope to be found in human ingenuity. Humans have a long history of finding new ways to tackle problems, and new ways to circumvent the worst. Without commitment from the highest levels, such ingenuity is likely to come to naught. But with such a commitment, and with a worldwide determination to make a serious cut in emissions, there is much that can usefully, and invigoratingly, be done. ■

Authorship policies

We are clarifying the duties of lead authors and making author-contribution statements mandatory.

The *Nature* journals encourage authors to treat their data and their collaborators with the respect that their communities demand. High-profile journals have a duty to reinforce the trends towards greater transparency and to help scientists to fulfil their responsibilities as researchers and authors. We are therefore introducing small but important changes in our policies to reflect these goals.

In a previous Editorial (*Nature* 450, 1; 2007), we asked for feedback about whether we should require senior or corresponding authors to sign a statement that they had taken some specific 'integrity insurance' steps before the manuscript was submitted. Some applauded this idea, but most were not in favour. (Some of the feedback can be seen at <http://tinyurl.com/ddjsxa>). Major doubts were expressed about the ability of corresponding authors to take on such responsibility given the diversity of collaborations. The belief was also expressed that such signed statements would too often be worthless box-ticking exercises. Although we regretfully accept these realities, we believe that we should go further in spelling out the responsibilities of co-authors, and in requiring an implicit acceptance of them.

Accordingly, we have modified the *Nature* journal policy on authorship, which is detailed on our website (<http://tinyurl.com/dkgbf8>). For papers submitted by collaborations, we now delineate the responsibilities of the senior members of each collaboration group on the paper. Before submitting the paper, at least one senior member from each collaborating group must take responsibility for their group's contribution. Three major responsibilities are covered: preservation of the original data on which the paper is based, verification that the figures and conclusions accurately reflect the data collected and that manipulations to images are in

accordance with *Nature* journal guidelines (<http://tinyurl.com/cmmrp7>), and minimization of obstacles to sharing materials, data and algorithms through appropriate planning.

Corresponding authors have multiple responsibilities, but we now make it clearer that the author list should include all appropriate researchers and no others, and that the order has been agreed to by all authors. They are expected to have notified all authors when the manuscript was submitted, they are the point of contact with the editor and they must communicate any matters that arise after publication to their co-authors.

Another change is that we have strengthened our policy for statements of authors' contributions. This policy was first introduced nearly 10 years ago (*Nature* 399, 393; 1999) to make the credit due to individual co-authors more explicit. Since then, authors of *Nature* papers have had the opportunity to include in their papers a statement that details each author's role in the published work. Over the past 10 years, the proportion of authors who choose to include this has risen dramatically.

This acceptance, and discussions with authors who have chosen not to include such a statement, has led us to change our policy. Rather than 'strongly encouraging' such statements, we now require them for publication of original research papers in *Nature* and the *Nature* research journals. The detail provided can vary tremendously (<http://tinyurl.com/39mmyw>) and authors are left to structure them as they see fit. We insist only that no author be left out.

To ensure that authors are familiar with these changes, we will shortly require the corresponding author to confirm that he or she has read the *Nature* journal policies on author responsibilities (<http://tinyurl.com/dkgbf8>) and is submitting the manuscript in accordance with those policies. ■

"We believe that we should go further in spelling out the responsibilities of co-authors, and in requiring an implicit acceptance of them."

RESEARCH HIGHLIGHTS

Dynamics of a dance

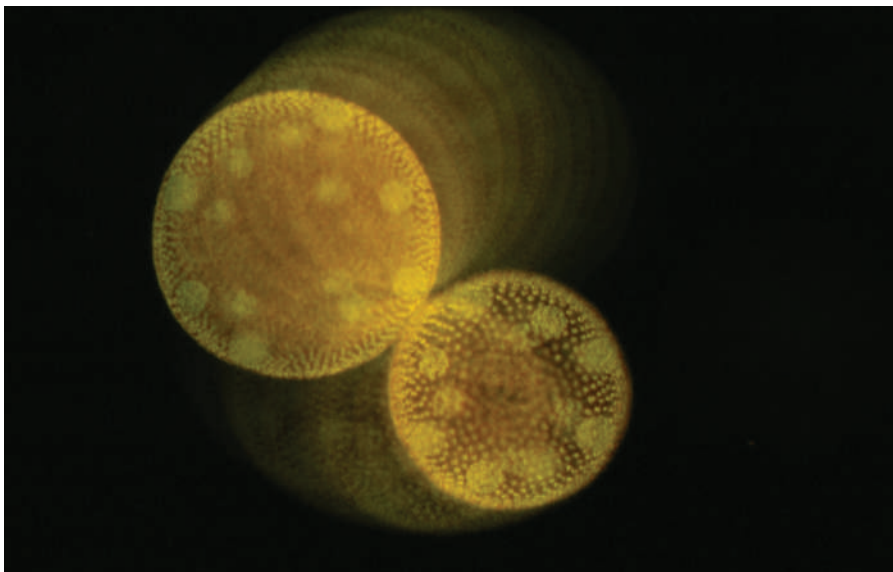
Phys. Rev. Lett. **102**, 168101 (2009)

Volvox are microscopic algae, spherical aggregates of thousands of flagellated cells. The spheres twirl about in ponds, swimming up and down in the water and interacting with their neighbours in seemingly orchestrated dances.

Physicist Raymond Goldstein of the University of Cambridge, UK, and his colleagues investigated the fluid dynamics used by *Volvox carteri* contained in a microscope-equipped chamber. Near the top of the chamber they observed the tiny dancers 'waltzing', spinning about each other in a clockwise fashion. Near the bottom, the spheres participated in a more complex 'minuet'.

The authors suggest that the careful dances increase female encounters with sperm packets during sexual reproduction.

For movies, see <http://tinyurl.com/degwdj>.



K. DRESCHER, R. E. GOLDSTEIN, UNIV. CAMBRIDGE

CHEMISTRY

Fuel from thin air

Angew. Chem. Int. Edn **48**, 3322–3325 (2009)

Carbon dioxide can be sucked out of the air (see News Feature, page 1094) and turned into a useful fuel using a metal-free catalyst.

The carbene catalyst — a compound with a pair of electrons available to react — is not only metal-free, and so better for the environment, it also works in air. Metal catalysts are often degraded by oxygen.

The system, which also uses a silicon-containing molecule, a silane, to activate carbon dioxide and drive the reaction, was developed by Jackie Ying, Yugen Zhang and Siti Nurhanna Riduan at the Institute of Bioengineering and Nanotechnology in Singapore.

The reaction product, methanol, can be turned into other carbon-based fuels, or used itself as a biofuel.

BIOCHEMISTRY

DNA base maker

Science doi:10.1126/science.1170116 (2009);

Science doi:10.1126/science.1169786 (2009)

Although four DNA bases — adenine, thymine, cytosine and guanine — make up much of the genome, modified bases can serve special purposes. Trypanosomes, parasitic protozoa, contain an additional

base called J that is a hypermodified version of thymine that has not been documented in other organisms.

Anjana Rao of Harvard Medical School and her colleagues searched for enzymes similar to those responsible for making base J that might make a similar base in mammals. They found TET1, which makes a modification to cytosine to create hydroxymethylcytosine. This accounts for 4–6% of all cytosines in the DNA of mouse embryonic stem cells.

Meanwhile, independently, Skirmantas Kriaucionis and Nathaniel Heintz at the Rockefeller University in New York have identified the modified base in the mouse brain.

STEM-CELL BIOLOGY

New stem-cell formula

Cell Stem Cell doi:10.1016/j.stem.2009.04.005 (2009)

Ever since Kyoto University's Shinya Yamanaka showed that cultured skin cells could be made to behave like embryonic stem cells by the addition of a handful of genes, researchers have been trying to repeat the trick without introducing DNA to the cells. Now, Sheng Ding at the Scripps Research Institute in La Jolla, California, and his colleagues say they can reprogram cells — in this case mouse embryonic fibroblasts — with those genes' protein products,

specifically engineered to cross the cellular and nuclear membranes.

The resulting cells are "morphologically indistinguishable" from embryonic stem cells, the authors say, and express similar markers.

Although such work has not yet been reported in human cells, Ding predicts that similar techniques will replace those requiring DNA or viruses, which are deemed risky in therapeutic applications.

For a longer story on this research, see

<http://tinyurl.com/cgonae>.

PHOTONICS

E-ink goes colour

Nature Photon. doi:10.1038/nphoton.2009.68 (2009)

Electronic readers without backlit screens can't beat the contrast and brightness of traditional ink and paper when it comes to colour.

Jason Heikenfeld of the University of Cincinnati, Ohio, and his colleagues, working with the Sun Chemical Corporation, also in Cincinnati, say that they have found a way to improve colour displays. Using inexpensive photolithographic techniques, the researchers made pixels with a reflective background and small wells containing water-dispersed pigments. Apply a voltage and pigment flows out of the well, coating the pixel (pictured left). Surface tension sucks the pigment back when voltage is removed; the switching is fast enough for video displays.

The researchers say their technique offers brightness, matt appearance and contrast that is superior to a related method that uses electric current to flip coloured oil droplets from beads to thin films across a pixel.



CHEMICAL BIOLOGY

Getting the glow

Nature Chem. Biol. doi:10.1038/nchembio.174 (2009)

Green fluorescent protein (GFP), originally isolated from the jellyfish *Aequorea victoria*, has had a transformative effect on biology. However, its purpose in nature has remained unclear. Now, Konstantin Lukyanov of the Shemyakin and Ovchinnikov Institute for Bioorganic Chemistry in Moscow and his colleagues have discovered that GFP can transfer electrons to certain proteins in a process powered by light.

When in the presence of certain electron acceptors such as cytochrome *c* or benzoquinone in an *in vitro* system, the authors noticed that signals changed from green to red. They suggest that the protein is passing an electron and changing conformation. They also found that GFP seems able to find protein electron acceptors in living cells.

Rather than being passive light-absorbing, glowing molecules, GFP may serve a chemical role, one that the authors suggest should be considered in the many research applications of this workhorse protein.

For a longer story on this research, see <http://tinyurl.com/dz8xrh>.

MATERIALS

Improving on nature

Science **324**, 488–492 (2009)

Spider silk is naturally tougher than steel, but adding metal makes it stronger still.

Seung-Mo Lee and Mato Knez of the Max Planck Institute of Microstructure Physics in Halle, Germany, and their colleagues took dragline silks from a spider caught in the institute gardens and pulsed them with metals in a process called multiple pulsed vapour-phase infiltration. Zinc oxide, titanium oxide or aluminium oxide not only coated the silk but also infiltrated the protein structure, resulting in much higher strength and extensibility.

The technique could be used on other biomaterials, the researchers say, such as collagen membranes from eggs.

GENOMICS

X-linked mysteries

Nature Genet. doi:10.1038/ng.367 (2009)

Efforts to resequence genetic variants that have been associated with disease may produce more questions than answers, at least at first.

Case in point: the results of Michael Stratton and Andrew Futreal of the Wellcome Trust Sanger Institute in Cambridge, UK, and their team, who have performed the largest resequencing study of its kind so far. They catalogued the protein-coding regions of 718 genes on the X chromosomes of individuals from 208 families affected by X-linked mental retardation.

By identifying variants predicted to truncate protein-coding genes and render them non-functional, the effort unearthed nine new genes probably involved in X-linked mental retardation. But it also found many similar truncating variants in normal individuals. The team estimates that people can function normally despite mutations that render 1–2% of genes on the X chromosome non-functional — a fact that could further complicate resequencing studies.



CLIMATE

Ground truths

Geophys. Res. Lett. doi:10.1029/2009GL037666 (2009)

Changes in land cover during the latter half of the twentieth century have affected the local climate and exacerbated droughts in eastern Australia. Replacing native vegetation with cropping or grazing lands can change how much sunlight is reflected and how much moisture evaporates, altering temperature and rainfall patterns.

Clive McAlpine of the University of Queensland in Australia and his colleagues simulated the period from 1951 to 2003 on computer climate models — comparing actual land use change with a scenario in which land stayed in its late-eighteenth-century pre-European state. The results imply that clearing native vegetation has worsened droughts, even if climate change is factored out.

BIOPHOTO ASSOCIATES/SPL

JOURNAL CLUB

Michelle Peckham
University of Leeds, UK

A cell biologist ponders an outstanding mystery in muscle formation.

Heart and skeletal muscles have a beautiful, almost crystalline structure of repeating contractile units called sarcomeres. The length of these units is precisely regulated along with the lengths of two types of overlapping filament (thick and thin) that they contain. Muscles contract when crossbridges from thick filaments interact with actin in thin filaments. The amount of contraction depends on the length of each filament and how much they overlap.

A thick filament contains exactly 294 myosin molecules — a limit imposed by the giant 'ruler' protein titin. Yet it is not clear what regulates the length of thin filaments. The protein nebulin has been a key candidate: its size corresponds to thin filament length in several species. Puzzlingly, however, in mice with a targeted deletion of nebulin, skeletal muscle thin filaments are the right length, at least at birth. And Ryan Littlefield at the University of Washington in Friday Harbor and his colleagues have now shown that nebulin is too short to be the ruler — its end is located just short of the tips of the thin filaments (A. Castillo *et al. Biophys. J.* **96**, 1856–1865; 2009).

Because of the way in which thick filaments are built, their middles — at the centre of the sarcomere — have no crossbridges. Littlefield and his colleagues suggest that thin filaments, which grow towards the middle of sarcomeres from their edges, stop growing when they reach this 'bare' zone. Intriguingly, this paper also shows that thin filament lengths in different muscles correspond to the length of titin in those muscles. A single titin molecule stretches from the edge to the middle of the sarcomere. If titin modulates overall sarcomere length, and thus the distance to the bare zone in the centre of the sarcomere, this could indirectly regulate thin filament lengths. Maybe the biggest protein known has yet another job.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Swine flu goes global

New influenza virus tests pandemic emergency preparedness.

Researchers are scrambling to study the evolution and spread of the novel H1N1 strain of swine influenza whose leap to humans was officially confirmed last week. The possible imminent onset of a swine-flu pandemic is also testing international preparedness plans put into place to deal with something else: the much-feared H5N1 avian flu virus that has spread across Asia, Europe and Africa since 2003.

The genetic make-up of this swine flu virus is unlike any that researchers have seen. It is an H1N1 strain that combines a triple assortment first identified in 1998 — including human, swine and avian influenza — with two new pig H3N2 virus genes from Eurasia, themselves of recent human origin.

“It has been mixing all over the place, and so the genetics are quite complicated,” says John McCauley, a virologist at the UK Medical Research Council’s National Institute for Medical Research in London. “Where the hell it got all these genes from we don’t know,” says Robert Webster, a flu virologist at St Jude Children’s Research Hospital in Memphis, Tennessee. “But this is a real super-mixed-up virus.”

As *Nature* went to press, more than 150 people were thought to have died from the virus in Mexico, with some 1,600 cases suspected there. At least 87 cases have been confirmed in other countries, including the United States, Canada, New Zealand, the United Kingdom, Spain and Israel.

On 27 April, the World Health Organization (WHO) raised its pandemic threat from phase 3 — a new flu virus infecting humans but with limited human-to-human spread — to phase 4, with community outbreaks that mark “a significant upwards shift in the risk for a pandemic”. The scale runs to 6: a full-blown pandemic.

“The scary thing is that this virus seems enormously transmissible,” says Webster. In New York, for instance, at least 28 students at a school in Queens have come down with it.

The US Centers for Disease Control and Prevention (CDC) and the WHO are trying to nail down how many of the reported deaths in Mexico are due to swine flu and how many are due to other causes. “The spectrum of illness is still a matter of doubt and conjecture,” says David Ozonoff, an epidemiologist at the Boston University School of Public Health in Massachusetts. For instance, rumours swirled that archaeologist Felipe Solis, the director of the National Museum of Anthropology in

Mexico City who had met recently with US President Barack Obama, died on 23 April as a result of the swine flu. Mexican authorities later reported he died of cardiac arrest linked to pneumonia.

“Where the hell it got all these genes from we don’t know.”

No human immunity

The virus seems to have first taken hold in Mexico in mid-March; it was identified when Mexican authorities sent samples from a flu patient that it could not subtype to the Canadian Public Health Agency. After cases in California, the CDC announced the existence of the virus on 23 April.

The genetics of the virus are so novel that humans are unlikely to have much immunity to it, scientists say. The current seasonal flu vaccine, which targets a different H1N1 strain, also isn’t likely to offer any protection. Discussions are under way as to whether a new vaccine for the swine flu strain should be produced. The WHO has recommended that vaccine makers continue to manufacture the seasonal flu strain but begin thinking about how to manufacture large doses of a vaccine that incorporates a weakened version of the



Face masks are being used — in Toluca, Mexico, and elsewhere — to try to stop the virus spreading.

M. TOVAR/AP PHOTO

current swine flu strain. For now, the virus is treatable with the influenza drugs oseltamivir (Tamiflu) and zanamivir (Relenza).

So far, cases in the United States and elsewhere seem to have been relatively mild compared with those in Mexico. Tashiro Masato, a virologist at the National Institute of Infectious Diseases in Tokyo, warns against drawing any conclusions from this. He notes that the sample of cases from the United States is too small to pick up even single-digit mortality rates. If transmission rates turn out to be high (see ‘Research questions to answer’), mortality rates of even a few per cent could lead to millions of deaths, as in the pandemic of 1918.

For the moment the pathogenicity and mortality range is wide and uncertain, says McCauley — “anything between the lethal 1918 and the mild 1968 pandemic”. The high transmission rates in areas such as New York are worrying, says Mark Lipsitch, an epidemiologist at the Harvard School of Public Health in Boston, Massachusetts. But a better idea should be forthcoming as testing ramps up, and health authorities get a clearer picture of the ratio of deaths to those infected.

The new virus, however, comes nowhere near the mortality rates of H5N1, which have averaged around 63% globally and reached as high as 82% in Indonesia. “The risk and threat of H5N1 remain as before”, Masato warns.

Epidemiologists are largely applauding the

RESEARCH QUESTIONS TO ANSWER

Virulence, including mortality rate: how the disease progresses within infected individuals.

Transmission characteristics: how quickly it spreads among individuals.

Genetics: how the virus acquired its jumble of genetic snippets.

Drug resistance: whether the new strain will evolve to be untreatable by the two flu drugs that can treat it now.

Vaccines: how to obtain and characterize viral isolates and decide which antigens to include in a vaccine.

Prevention: how to limit spread and buy time until a vaccine is available. **D.B.**



speed and scale of the response to the swine flu outbreak since it was formally identified last week, even though Mexican and international disease surveillance systems failed to pick it up promptly at the start. Several scientists speculate that the initial cases were months ago and perhaps outside Mexico. Ambitious WHO plans to quickly detect and extinguish an emerging pandemic before it grew, by distributing flu drugs, were thus not useful.

But once the virus was officially announced, governments have swung into action. The United States, for instance, has released one-quarter of its 50 million stockpiled doses of oseltamivir and zanamivir for treatment in the affected states. The international response is being largely attributed to the amount of pandemic planning undertaken since the threat of a H5N1 pandemic emerged. Gene sequences of the virus samples, for instance, have been promptly shared on the Global Initiative on Sharing Avian Influenza Data (GISAID) database. Sequence information can have practical significance — for example, in designing PCR primers, to make rapid diagnosis tests that broaden and speed up surveillance.

Many virologists think they are in for the long haul. “Will this fizzle out? It doesn’t look like it to me,” says McCauley. “For me this is the time to start deploying national pandemic plans. We need to act now.” ■

Declan Butler

For Nature’s coverage of the swine flu outbreak, see www.nature.com/swineflu.



MOON DUST

‘Lost’ Apollo tapes help plan future lunar missions.

www.nature.com/news

NASA

California in clean-fuel drive

The state of California has adopted regulations to curb greenhouse-gas emissions from transportation fuels, codifying evidence that biofuels are significantly dirtier than they were once thought to be.

The California Air Resources Board approved its ‘low-carbon fuel standard’ on 23 April, requiring fuel providers to cut the greenhouse-gas emissions from fuels by 10% by 2020, compared with 2010 levels. The rule would also accomplish 10% of California’s requirement to reduce emissions to 1990 levels by 2020, part of a broader plan to cut emissions by 80% by mid-century.

“We really think this is the future of fuels policy globally,” says Anthony Eggert, a science adviser to the board. He says the policy should send a signal to the private sector and “drive clean transportation fuels and infrastructure into the market”.

The regulation includes greenhouse-gas-emissions calculations for the life cycle of all fuels. To meet the requirement, refiners and importers would have to improve their production methods or mix and match the fuels that they sell. If they beat the standard, they can sell credits for their unused emissions to someone else.

Despite intense opposition from the US corn (maize) ethanol industry, the rule takes into account agricultural expansion abroad caused by rising grain prices as food crops are diverted for biofuels. This ‘indirect’ effect boosts the estimated emissions for various categories of corn ethanol by 50% or more, meaning that ethanol often results in higher greenhouse-gas emissions than gasoline.

Matt Hartwig, a spokesman for the Renewable Fuels Association in Washington DC, questions the science behind indirect emissions and says California has vastly overstated the problem. Many researchers, however, think there is enough evidence to move forwards now and re-evaluate as the science improves.

Dan Kammen, an energy researcher at the University of California, Berkeley,

says California’s calculation for indirect emissions is “reasonably consistent” with his team’s work. “The science is new and evolving,” he says, “but the low-carbon fuel standard is exactly the right approach.”

The US Environmental Protection Agency is reviewing its own ruling that will establish greenhouse-gas criteria under the national biofuels mandate. European regulators are also looking into indirect emissions, but are not expected to make a determination until next year.

Tim Searchinger, an environmental researcher at Princeton University in New Jersey, identified the indirect effect in 2008 (T. Searchinger *et al. Science* 319, 1238–1240; 2008). He says California was conservative in its calculation of indirect emissions. His work suggests that corn ethanol could double emissions compared with gasoline over 30 years.

But one of the reasons for this difference, he says, is because California assumes that half of the food crops that are redirected towards energy will not be replaced because of rising prices, which he says translates into hunger in poorer countries. “Even though it’s a best-case scenario for corn ethanol, the impact is still significant,” Searchinger says. “Corn ethanol basically has no benefit, and it causes hunger.”

Brazilian sugarcane ethanol, despite

having an even higher indirect effect and being transported abroad, still performs better than any other biofuel. The California regulation does not yet include specific numbers for biodiesel, although the indirect impact could be even higher because of massive

emissions from expanding palm-oil plantations, particularly in Indonesia.

The California regulation also covers fuels produced from the Canadian tar sands, which produce more emissions than conventional oil because of the energy-intensive extraction process. The low-carbon fuel standard could also prove a barrier to coal-based fuels, while providing a boost to vehicles powered by natural gas and electricity.

Jeff Tollefson



California dreaming of much less smoggy days.

K. DIANSEZIAN/AP PHOTO

Japan goes for the sun

Tens of thousands of Japanese homes and businesses are preparing to put new solar panels on their roofs, spurred by a subsidy plan going into effect this month. It is a step towards regaining Japan's once-dominant position in solar energy, something that has become a matter of pride for the country's policy-makers.

But the government knows it cannot count on citizens to continue backing solar energy out of civic pride and eco-mindedness alone, so it is now spending ¥30 billion (US\$300 million) of state money annually on subsidies and research and development on the technology.

In 1999 Japan led the world in solar-cell production, a position it maintained for several years. In 2005, solar cells produced by Japan accounted for 45% of the world's capacity, according to figures from the country's industry ministry. By 2007, however, the figure was only 24.6%, barely ahead of China's 22.0%.

Each year from the late 1990s on, Japan also installed more capacity for producing energy from solar cells than any other country, but was surpassed by Germany in 2005 and Spain in 2008. It more than doubled its total capacity between 2003 and 2007, but could not keep pace with Germany's ninefold increase over the same period. The German boom was fuelled by



RECESSION
WATCH

"Germans do it because it is profitable, but there's no profit for the Japanese."

a 'feed-in' tariff which guarantees that solar-cell power plants and rooftop installations receive a high price for electricity that they feed back into the grid. By 2007, Germany had twice Japan's 1.92 gigawatts of installed capacity.

Japan now wants to get back on top. Last July, the cabinet announced a low-carbon-economy action plan that called for the country to "regain its global leadership position in solar generation". It set out ambitious targets to increase its 1.4 gigawatts of solar power capacity in 2005 20-fold by 2020. The 2009 budget earmarks ¥10 billion for solar-photovoltaic energy research and development. Another ¥60 billion has been set aside for renewable energy subsidies, a third of which will go

on solar cells.

The country's recently announced economic stimulus package also targets solar research and development as a major investment area, although specific numbers have not yet been determined (*Nature* 458, 819; 2009).

The subsidy programme starting this month gives ¥70,000 per kilowatt to homes and businesses that install solar panels. So far it is proving popular, with 22,000 applications in its first two-and-a-half months.

Takashi Kawabata of the industry ministry's



new and renewable energy division notes that Japan is unusual in having so much of its capacity — 1.55 million kilowatt-hours, or 80% — in homes and businesses. In Germany, only 40% of the country's capacity is found in these places; instead, power companies have found it profitable to invest because of the feed-in tariffs. "Germans do it because it is profitable," Kawabata says, "but there's no profit for the Japanese." To explain his fellow-citizens' motivation for installing the technology, he cites

Y. NAKAO/REUTERS

Basic researchers protest UK budget

Britain's government has unveiled an economic stimulus package designed to harness what it calls a "world-class science base" — at the same time as it cuts funds for undirected basic research.

In a speech to parliament on 22 April, Alistair Darling, the Chancellor of the Exchequer, unveiled the 2009 budget. It includes £1.4 billion (US\$2.1 billion) in new cash for low-carbon business and technology.

Some scientists were underwhelmed by the plan, which redirects £106 million from the nation's seven research councils,

which fund most of Britain's basic research, towards "key areas of economic potential". These areas make up roughly 15% of the councils' £3-billion annual budget and are defined by five cross-cutting programmes with titles such as "living with environmental change" and "digital economy" (see *Nature* 453, 1150–1151; 2008).

Chloë Somers, a spokeswoman for the councils, says that most of the £106 million will come from reprioritizing the 'blue-skies'

funding that allows researchers to pursue any topic they choose. Despite the change in focus, Somers argues, the councils will still fund fundamental science: "This has

"The value of our universities lies in their transformative discoveries."

nothing to do with cutting basic-research funding."

But others see it as an erosion of the councils'

independence. The decision is part of a worrying trend to force basic research to prove its economic worth, says Nick Dusic, director of the Campaign for Science and Engineering in the UK, a London-

based group that advocates for science funding. If it continues, it could undermine the real worth of basic science in Britain, adds Martin Rees, president of the Royal Society. "The value of our universities lies in the transformative discoveries that emerge unpredictably and unplanned," he says. "The research councils should not stifle this potential."

Environmentalists, meanwhile, were cheered by the government's green stimulus package, which includes £1 billion to support renewable energy and jobs in green businesses. The development



Japan knows how to do solar power at home.

the well-known Japanese word *yasegaman*, which means 'endure for the sake of pride'.

Nevertheless, Japan is debating a feed-in tariff of its own, which could come into effect in 2010 and would allow homes and businesses with solar panels to sell electricity back to the grid at twice the going rate. Many worry, however, that it would force electricity prices up overall. In Japan, the cost of producing a

kilowatt-hour of electricity using solar cells is twice that charged to consumers.

The government aims to cut that cost in half within five years. The New Energy and Industrial Technology Development Organization, which distributes research money for the industry ministry, has been leading an eight-year field test to prove the value of large-scale solar panel installations. It is spending roughly ¥6.3 billion per year from 2007 to 2014 on this project, and about another ¥3.7 billion for various others aimed at raising solar-cell efficiency, making manufacturing processes cheaper and finding ways to use cheaper materials.

In one of these projects, some 30 university groups and companies have banded together for solar-cell research. Mitsubishi Electric, for example, has made polycrystalline solar cells that convert 18.8% of solar energy to electricity, compared with the 15–16% achieved by most on the market now. Sanyo's improved single-crystal cell has 22.3% efficiency — the world's best, according to project leader Masafumi Yamaguchi, a semiconductor materials scientist at the Toyota Technological Institute in Nagoya.

Yamaguchi is also the research adviser for a new programme funded by the Japan Science and Technology Agency. With a hefty ¥8 billion over seven years, 12 teams of scientists will try to find new materials to make solar cells cheaper.

Yamaguchi says Japan needs to pursue this course because of the pressure from cheap producers, especially in China. "It's the only way to compete," he says.

David Cyranoski



Q&A: LARRY BRILLIANT
Google's former philanthropy chief on saving the planet.
www.nature.com/news

Obama says more money

President Barack Obama has bolstered his already-lofty promises to US scientists, saying he wants the country's research and development budget to rise to 3% of gross domestic product (GDP).

That would be an increase of roughly US\$46 billion annually from today's investment, which is nearly 2.7% of GDP. "This represents the largest commitment to scientific research and innovation in American history," Obama said on 27 April at the National Academy of Sciences' annual meeting in Washington DC.

John Marburger, who was science adviser to former president George W. Bush, called 3% of GDP a "healthy target" but said the trick will be getting industry on board. "The federal government can't do all of that by itself," he says. "Remember, two-thirds of that figure is coming from the private sector, and we're in the middle of a recession." The public sector contributes roughly one-third of US research and development investment.

"We have to deliver."

It remains to be seen whether Obama can

deliver. In February, he and congressional leaders pushed through over \$21 billion in science funding in their economic stimulus package. In his proposed budget for the fiscal year that begins in October, the president has requested \$12.6 billion for the National Science Foundation, the energy department's Office of Science and the National Institute of Standards and Technology.

Obama called clean energy the current generation's "great project" and said that investment levels must be increased despite ongoing economic woes.

"We're now in a no-excuse environment," says J. Craig Venter, the genomics pioneer now working on sustainable energy issues. "We have to deliver."

This week Obama also named all 20 members of the President's Council of Advisers on Science and Technology, including Nobel-prizewinning chemist Ahmed Zewail, computational scientist David Shaw, plant geneticist Barbara Schaal and physicist S. James Gates.

Jeff Tollefson

S. DAWSON/BLOOMBERG NEWS/LANDOV/PA PHOTOS



Alistair Darling announced the UK budget on 22 April.

of offshore wind farms is expected to receive £525 million over the next two years. The budget also

includes a new £750-million strategic investment fund for emerging technologies and high-tech industry and

sets aside £405 million to support low-carbon energy and advanced green manufacturing. Biotechnology will also receive some of the money.

Darling told parliament that together, the incentives will "harness commercially our world-class science base".

That may be true, but the base is being ignored, Dusic argues. The United States is using stimulus money to make mammoth investments in basic science, he says, so "the United Kingdom cannot afford to fall behind".

Geoff Brumfiel and Natasha Gilbert

Funding cut for US nuclear waste dump

Yucca Mountain's end would leave the country with few alternatives for a long-term repository.

Some ten billion dollars and two decades into the project, the proposed US nuclear waste dump at Nevada's Yucca Mountain has hit a major and possibly decisive stumbling block: President Barack Obama has proposed eliminating all funding for scientific research on the deep-rock repository, 140 kilometres northwest of Las Vegas. As Yucca Mountain has been the United States' only potential long-term repository since 1987, the decision once again raises the issue of what to do with the country's high-level nuclear waste.

In his budget for fiscal year 2009, to be released in early May, Obama is expected to propose spending only the estimated US\$40 million needed to let the Nuclear Regulatory Commission (NRC) move forwards in evaluating the licence application for Yucca Mountain. The licence request, filed last June under George W. Bush's administration, requires the NRC to investigate the site's technical suitability, a process expected to take three to four years.

But the new administration says it will halt the controversial project regardless of the licence ruling, saying that the country can find another solution for long-term management of its nuclear waste. "I think we can do better," energy secretary Steven Chu told the Senate committee on energy and natural resources in March. The Department of Energy has established a commission to explore alternatives, which could report back within a year.

For now, high-level waste remains in cooling ponds and dry-cask storage containers at the country's 104 nuclear power plants. But that is a short-term and expensive solution: since the government missed a 1998 deadline for opening the Yucca Mountain facility, it has been paying fees to the power plants, a bill that could rise to \$11 billion by 2017. Some states have threatened to stop contributing unless an alternative to Yucca Mountain is found.

Tom Kauffman, a spokesman for the industry policy group the Nuclear Energy Institute, says that waste can be safely stored on-site "indefinitely" while a long-term solution is found, but others say an answer is needed. "I live in a state with three [nuclear] power plants," says Michael Corradini, a nuclear engineer at the University of Wisconsin in Madison. "It's very unsatisfying to not have a complete, thought-out process."

One possibility is to begin searching for another geologic repository. Yucca Mountain's capacity was to be legally capped at 70,000



US DEPARTMENT OF ENERGY/SPL

Billions of dollars spent on the proposed Yucca Mountain repository could be wasted.

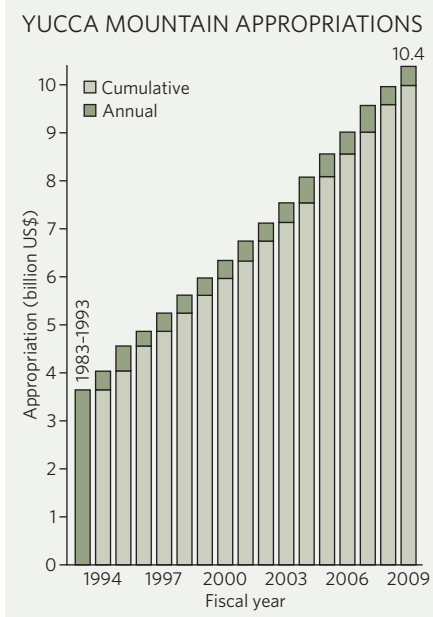
tonnes of nuclear waste, even though it could technically hold 4–9 times that amount, so the US government had intended to open a second site, says Charles McCombie, a nuclear-waste consultant based in Gipf-Oberfrick, Switzerland. "It would be sensible for the United States to begin looking at another repository-siting

programme, whether or not Yucca Mountain shapes up," he says.

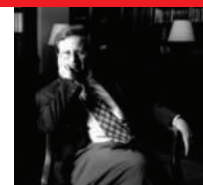
But the government has spent decades studying Yucca Mountain as the only potential site, spending more than \$10 billion (see chart) to evaluate its geologic suitability. And few are enthusiastic about restarting the bitter political arguments over where to put a repository. Sites in Washington state and Texas, which were under consideration until Nevada was controversially selected, might be options, as is an area in southeast New Mexico near the Department of Energy's Waste Isolation Pilot Plant, which stores military nuclear waste and where the local community has expressed interest in hosting a site.

Starting anew

The political challenges surrounding Yucca Mountain might prompt the Obama administration to take a fresh look at reprocessing: purifying plutonium-239 and uranium-235 and recycling those fuels to reduce the waste's mass and volume. Because the pure materials can be used in nuclear weapons, the US suspended reprocessing of spent fuel in 1977. The previous administration began a programme, the Global Nuclear Energy Partnership, to explore an international reprocessing project, but Congress gave the programme less than half the money



SOURCE: OCRWM, NYTIMES



ENERGY-SCIENCE CZAR
Steve Koonin heading for
US Department of Energy
post.
www.nature.com/news

CALTECH

that Bush requested. Under Obama, the project may stop altogether.

"Leaving it where it is, putting it centrally or reprocessing it can all be done. But in all cases, there has to be a final resting point for the solid waste," says Corradini. "It's pretty clear from an international perspective that geologic disposal of high-level waste, and probably of any nuclear waste, is the favoured option," adds John Garrick, chairman of the US Nuclear Waste Technical Review Board, based in Arlington, Virginia.

Some countries are well on their way towards siting a repository: Finland is working on a repository at Olkiluoto, which could be operational by 2020. Sweden is finalizing research on two candidate disposal sites and expects to choose one by the end of the year.

In both Finland and Sweden, the key to success was that governments adopted a cautious time schedule and a stepwise approach and involved the local community from the start, says Barbara Pastina, a nuclear-waste scientist at the Helsinki-based engineering firm Saanio and Riekkola Oy. In Finland, she says, "the whole nation was represented in the decision to build the repository" at Olkiluoto. The community had the right to veto the location at the beginning, and parliament voted on the decision. After early setbacks, Canada and the United Kingdom are also on track to find geologic repositories for long-term storage. Canada reorganized its national waste-management efforts in 2002 and is moving towards a siting phase, whereas the United Kingdom restarted in 2001.

By contrast, powerful states' efforts to avoid hosting the repository have stymied US attempts to find a site. Senator Harry Reid (Democrat, Nevada), now the Senate majority leader, has fiercely opposed the project from the beginning, and many Nevadans are still fighting the Yucca Mountain proposal. "They've always felt that it was hoisted on them rather than them volunteering," says Todd Allen, an assistant professor of nuclear engineering at the University of Wisconsin.

Public consent has been the missing ingredient in Yucca Mountain from the beginning, agrees McCombie. The trick to finding another location, he says, is a "modern approach that combines good science with a consensus of enough people that it's the right thing to do." The United States hasn't built a nuclear power plant in three decades, but a resurgence of nuclear power is on the horizon. The NRC is considering licences for 26 new nuclear power plants, and construction could begin as early as 2012 — increasing the pressure on the country to solve long-term disposal once and for all. ■

Amanda Leigh Mascarelli

Brain imaging skewed

Double dipping of data magnifies errors in functional MRI scans.

Nearly half of the neuroimaging studies published in prestige journals in 2008 contain unintentionally biased data that could distort their scientific conclusions, according to scientists at the National Institute of Mental Health in Bethesda, Maryland.

Experts in the field contacted by *Nature* have been taken aback by the extent of the methodological errors getting through the supposedly strict peer-review systems of the journals in question.

Nikolaus Kriegeskorte, Chris Baker and their colleagues analysed 134 functional magnetic resonance imaging (fMRI) studies published last year in five top journals —

Nature, *Science*, *Nature Neuroscience*, *Neuron* and *The Journal of Neuroscience*.

The survey, published in *Nature Neuroscience* on 26 April (N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan and C. I. Baker

Nature Neurosci. 12, 535–540; 2009), found that 57 of these papers included at least one so-called 'non-independent selective analysis'; another 20 may also have done so, but did not provide enough information to confirm suspicions.

The non-independence of the analysis lies in using the same data to set up the conditions to test a hypothesis, then to confirm it. "We are not saying that the papers draw wrong conclusions, because in some cases the error will not have been critical," says Baker. "But in other cases we don't know, and this creates an ambiguity."

"It is a poor reflection on the quality of peer review of prestige journals — they really need to up their game in terms of rigour," says Karl Friston, scientific director of the Wellcome Trust Centre for Neuroimaging at University College London.

Brain imaging provides vast quantities of data in the form of voxels (three-dimensional pixels), from the entire brain. Neuroscientists sometimes focus on an area of interest by searching for voxels that are activated when subjects perform different tasks in an experiment — for example, looking at a face or an inanimate object.

But fMRI data are intrinsically very noisy, producing many 'false voxels'. Problems arise when researchers use the same data to select a particular brain region and then to

quantify the experimental effects there — for example, by asking how much more strongly the region responds to a face compared with an inanimate object.

"It is crucial to analyse your results with a set of data that are independent of that used in the earlier selection process," says Chris Baker. "It is even OK to split your total data and use one half to select the voxels, and the other to further analyse the response in these voxels."

A similar type of error has been addressed by Edward Vul of the Massachusetts Institute of Technology in Cambridge and his colleagues (E. Vul, C. Harris, P. Winkielman and H. Pashler, *Perspect. Psychol. Sci.* 4,

274–290; 2009). A preprint of their research caused uproar in the field earlier this year by referring to 'voodoo correlations' and naming labs it considered guilty of circular analysis (see *Nature* 457, 245; 2009).

In contrast, the study by Kriegeskorte and Baker does not single out any researchers. "We didn't name names because the error is just too common," says Baker. "And we saw no reason to be personal — our idea was to highlight a problem so people are less likely to fall into the trap."

"This new paper is less controversial, but potentially more worrying," says Friston. "The issue of selection bias doesn't require special understanding of statistics, just the following of good practice — it is not rocket science."

Baker points out that circularity errors creep into many areas of neuroscience. "It applies equally to single-unit electrophysiology, electroencephalography, gene microarray studies or even behavioural data," he says. But fMRI data are particularly vulnerable because of the complex analysis demanded by their huge volumes, and because so many untrained outsiders are entering the field. "For those of us with a few years of fMRI experience the issue is entirely passé, but there will always be a substantial minority on a steep learning curve," says Friston. "What surprised me is how frequent the errors are."

Baker notes that the increasing complexity of the data "probably leads people to take their eye off the ball so that the more fundamental aspects don't get taken care of". ■

Alison Abbott

"The issue of selection bias doesn't require special understanding of statistics — it is not rocket science."



SETBACK FOR ROCHE
Avastin fails in colon cancer trials.
www.nature.com/news

BSIP, L. OLIGNY/SPL

Fake Facebook pages spin web of deceit

Stem-cell scientists are caught up in fictional friend network — but no-one knows why.

In September 2008, *Forbes* science editor Matthew Herper and former *Washington Post* reporter Rick Weiss appeared together on a panel at the World Stem Cell Summit in Madison, Wisconsin. In late February, Herper received an invitation to 'friend' Weiss on the Internet social-networking site Facebook. On the basis of their acquaintance, Herper accepted, noticing that a number of other people involved with stem cells were listed as friends on Weiss's profile. However, that profile — and many of those it was linked to — was a fake.

Weiss is one of more than 100 scientists, policy-makers and journalists, many linked to stem-cell research, whose identities have been purloined to create a convincing — but bogus — network of apparent friends on Facebook. The victims have no control over the profiles that carry their names, and the perpetrators — and their motive — remain unknown.

Security experts call the creation of a network of impostors a Sybil attack. "People use impostors to create the image that a certain idea is more accepted or under more fire than it really is," says John Wilbanks, vice-president of science at the non-profit corporation Creative Commons, which provides free intellectual-property licences for online content.

The second of those possibilities was a worry to Alta Charo, a bioethicist at the University of Wisconsin-Madison, when she discovered fake Facebook profiles in her name and in those of other Wisconsin researchers with connections to stem-cell work in February. The list included Carolyn 'Biddy' Martin, who attended the stem-cell summit in her role as university chancellor. Facebook removed the profiles after the university's legal counsel contacted the company. A Facebook official told attendees at the e-Crime Congress in London last month that the company removes thousands of profiles a week.

Charo has no real Facebook profile: the fake one contained information easily gleaned from web searches as well as some made-up but harmless material. Weiss did have a real



Matthew Herper (top) and Rick Weiss were targets of a 'Sybil' attack.

"Impostors create the image that a certain idea is more accepted or under more fire than it really is."

profile under his own control, and updates to this were channelled to the doppelgänger, adding verisimilitude. The bogus site suggests that at least one of the true Weiss's friends was an impostor. What it lacks, according to someone who has seen both profiles, is the affiliation that links Weiss's real profile to the network of the think-tank the Center for American Progress, where Weiss worked before becoming the communications director and policy analyst at the White House Office of Science & Technology Policy. Such affiliations require e-mail confirmation, and none of the bogus profiles has any.

Davide Balzarotti, a computer-security researcher at the communication systems research centre Eurecom in Sophia-Antipolis, France, estimates that it would take only a day or so for a programmer to create the stem-cell

impostor network using a mixture of manual and automated processes to harvest data from elsewhere. Balzarotti and colleagues have created cloning software that replicates networks found on one site on a second — such as LinkedIn, XING and Facebook — in order to test the ease with which a Sybil attack could be automated (L. Bilge *et al.* 'All your contacts are belong to us: automated identity theft attacks on social networks.' WWW 2009 conf., Madrid, 20–24 April 2009). Balzarotti's Eurecom colleague Engin Kirda sees nothing to suggest that the scientists were targeted for imposture because of their specific work, but says it is possible to imagine a researcher creating a false profile and a Sybil network around it in order to obtain sensitive information from a hoodwinked competitor.

There is little activity on the impostor network, but it is not moribund: two of the suspected impostor profiles disappeared from public view after *Nature* tried to make contact (we have also brought the situation to Facebook's attention). And the false Rick Weiss profile has had one comment on it from a 'John Birch' that didn't appear on the real profile. John Birch, who died in China in 1945, is the namesake of a right-wing political organization, the John Birch Society, which is headquartered in Appleton, Wisconsin. The society, like many on the American right, is against research that involves viable human embryonic stem cells, says William Jasper, senior editor of the society's magazine, *The New American*. He had not heard of the impostor network, but noted that John Birch is a common name that could easily be real or fake.

Although no harm seems to have been done, the experience of being cloned remains a disconcerting one. As Elaine Fuchs, a stem-cell researcher at Rockefeller University in New York, puts it: "I'm afraid to open it ... I feel safer remaining completely oblivious to whatever is in there." Instead, she asked an official from her institution to deal with it. She writes, "I guess I'm not really interested in trying to come up with speculation about the mind(lessness) of the person who did".

Lucas Laursen

2008 WORLD STEM CELL SUMMIT

Close shave for Austrian science budget

The Austrian government has retreated from its threat to cut the science budget by 40% (see *Nature* 457, 648; 2009), approving instead a small rise that slows rather than reverses the country's plans for a big expansion in science.

The FWF, Austria's principal research-funding agency, has had its money cut by 18%, but president Christoph Kratky says the new budget stability makes up for the shortfall. In the next few weeks the FWF will clear a backlog of 700 reviewed projects accumulated over six months of uncertainty.

The Austrian Academy of Sciences, which runs 33 research institutes, gets a bare 2% rise, although it has recently opened several new institutes. Outgoing academy president Peter Schuster says that these — including the Centre for Molecular Medicine and the Institute of Molecular Biotechnology, both in Vienna — will be sheltered from cuts. The academy is now deciding where to wield the knife among the older institutes.

For a longer version of this story, see <http://tinyurl.com/cnahq3>.

Nobel laureate celebrates her centenary

The Italian neuroscientist Rita Levi-Montalcini became the first Nobel laureate to reach the age of 100, on 22 April. During birthday celebrations at a scientific meeting in Rome, Italian education minister Maria Stella Gelmini announced two gifts to honour her.

One was a €500,000 (US\$650,000) one-off sum to the research institute that Levi-Montalcini founded, the European Brain Research Institute in Rome, which is close to bankruptcy (see *Nature* 458, 564–567; 2009). The second was a €6-million programme, named after her, to return 30 young scientists working abroad to Italy on three-year grants.

Many scientists criticized the donations as drops in the ocean, but Levi-Montalcini said the gestures were more than she had dared



Rita Levi-Montalcini: first Nobel centenary.

Researchers rally to support animal studies

Advocates for research outnumbered animal-rights activists by ten to one at a peaceful demonstration last week at the University of California, Los Angeles (UCLA).

UCLA Pro-Test, a group formed to support scientists who use animals in their studies, rallied about 400 people on 22 April — an estimated 40 animal activists also attended.

Pro-Test was formed by UCLA neuroscientist David Jentsch, whose car was firebombed in March by animal extremists reacting to recent arrests (see *Nature* 458, 813; 2009).



REED SAXON/AP

to imagine. She did not seem to tire during nearly a week of public celebrations.

Japan cuts red tape holding up stem-cell work

A Japanese science ministry committee announced a plan on 22 April to loosen regulations on research involving embryonic stem cells (ES cells).

Scientists have complained about the current excessively bureaucratic system whereby they must seek permission for such work from a ministerial commission. The commission meets only rarely, and sometimes requests extra data from animal experiments to justify human research.

Under the new plan, expected to be approved by the Council for Science and Technology Policy, ES-cell researchers will only have to gain approval from their local institutional review board.

From May, the ministry will also free up the previously forbidden creation of cloned human embryos, although only for basic research into serious diseases.

New UK coal must be partly 'clean'

Any new coal-fired power station built in Britain must deploy carbon capture and storage (CCS) technology on 400 megawatts of its output, the country's energy and climate minister Ed Miliband announced on 23 April. If the Environment Agency judges CCS to be technically and economically 'proven', it would have to be retrofitted to cover the full output of new plants (typically 1,000–2,000 megawatts) by 2025.

Miliband's speech followed the government's 22 April budget commitment to fund up to four CCS demonstrations. A levy raised from electricity suppliers — and ultimately consumers' bills — would raise

the billions of pounds needed to support the projects, the Department of Energy and Climate Change said. A consultation this summer will firm up these plans.

The budget (see page 1084) also committed Britain to a cut in greenhouse-gas emissions of 34% from 1990 levels by 2020, a milestone towards meeting the government's stated goal of an 80% reduction in emissions by 2050.

Texas agencies sue over national disease lab

A consortium of Texas agencies that lost a bid for the new US\$560-million-plus US National Bio and Agro-Defense Facility has sued in federal court to overturn the government's choice.

Earlier this year, the Department of Homeland Security finalized the selection of Kansas State University in Manhattan, Kansas, to host high-containment laboratories that will replace the ageing Plum Island Animal Disease Center in New York state (see *Nature* 456, 687; 2008).

On 23 April, the Texas Biological and Agro-Defense Consortium in San Antonio sued the department, alleging faulty decisions and political manipulation during the three-year competition process.

The consortium alleges that the department under-secretary who made the decision, Jay Cohen, was biased towards Kansas, and that the Kansas location is unsafe because of tornadoes and the cattle herds that could contract disease from an accidental release. A hearing is set for 8 June.

Correction

The News Feature 'The test of inflation' (*Nature* 458, 820–824; 2009) misstated the implications of Planck's hunt for B-modes. The sentence on page 824 should have read: 'If Planck fails to detect a B-mode signal, but does find evidence for non-Gaussianity, inflation would certainly be "in more of a corner", according to Turok.'



A BURDEN BEYOND BEARING

The climate situation may be even worse than you think. In the first of three features, **Richard Monastersky** looks at evidence that keeping carbon dioxide beneath dangerous levels is tougher than previously thought.

In 2007, environmental writer Bill McKibben approached climate scientist James Hansen and asked him what atmospheric concentration of carbon dioxide could be considered safe. Hansen's reaction: "I don't know, but I'll get back to you."

After he had mulled it over, Hansen started to suspect that he and many other scientists had underestimated the long-term effects of greenhouse warming. Atmospheric concentration of CO₂ at the time was rising past 382 parts per million (p.p.m.), a full 100 ticks above its pre-industrial level. Most researchers, including Hansen, had been focusing on 450 p.p.m. as a target that would avoid, in the resonant and legally binding formulation of the United Nations Framework Convention on Climate Change, "dangerous climate change". McKibben was aware of this: he was thinking of forming an organization called 450.org to call attention to the number, and his question to Hansen was by way of due diligence.

As he thought about McKibben's question, Hansen, who runs NASA's Goddard Institute for Space Studies in New York, began to wonder if 450 p.p.m. was too high. Having spent his career working on climate models, he was



aware that in some respects the real world was outstripping them. Arctic sea ice was reaching record lows; many of Greenland's glaciers were retreating; the tropics were expanding. "What was clear was that climate models are our weakest tool, in that you can't trust their sensitivity in any of these key areas," he says. Those warning signs — and his studies of past climate change — led Hansen to conclude that only by pulling CO₂ concentrations down below today's value could humanity avert serious problems. He came back to McKibben with not 450 but 350. In 2008, he published a paper spelling out his rationale for that target¹.

The difference between 350 and 450 is not just one of degree. It's one of direction. A CO₂ concentration of 450 p.p.m. awaits the world at some point in the future that might conceivably, though with difficulty, be averted. But 350 p.p.m. can be seen only in the rear-view mirror. Hansen believes that CO₂ levels already exceed those that would provide long-term safety, and the world needs not just to stop but to reverse course. Although his view is far from universal, a growing number of scientists agree that the CO₂ challenge is even

greater than had previously been thought.

Several recent studies, for example, indicate that it may be exceedingly difficult to cool the climate down from any eventual peak or plateau, no matter what CO₂ concentration is chosen as a target by the international community. And by looking at the problem in a new sort of way — by tallying the total amount of carbon injected into the atmosphere across human history — two papers in this issue of *Nature* reveal how close the world has come to the danger point (pages 1158 and 1163). "It's tougher than people have appreciated. We have less room to manoeuvre," says Malte Meinshausen, an author of one of the papers and a senior researcher at the Potsdam Institute for Climate Impact Research in Germany.

Mr Greenhouse

Hansen has a long history of stirring up controversy with gloomy climate prognostications. Often, they turn out to be right. In 1988, he told the US Congress that the recent warming of Earth's surface was very unusual and it was time to point a finger at the cause. Hansen said it was his opinion that "the greenhouse effect has been detected and it is changing our climate now". He caught a lot of flak for that

ILLUSTRATIONS BY J. BURTON

statement, but the Earth continued to heat up and the rest of the scientific community eventually concurred with his assessment. He also used models to predict the amount of subsequent cooling to be expected from the eruption of Mount Pinatubo in 1991. That did much to convince people of the reliability of such models and of climate theory.

The model simulations Hansen and others worked on in the 1970s and 1980s had a profound effect on both climate scientists and politicians. When nations started exploring policies to curb CO₂ emissions, the target most discussed was 550 p.p.m., in large part simply because that was what the modellers had experience with: in early studies of the greenhouse future, researchers had sought to get a sense of the scale of possible change by simulating what would happen if the atmosphere held 550 p.p.m., roughly twice the pre-industrial level of CO₂ in the air.

Those studies showed a 550-p.p.m. world as warming quite a lot. In 1979, a panel of the US National Academy of Sciences led by Jule Charney, a prominent weather and climate researcher, estimated it would be 1.5 to 4.5°C hotter. That estimate for what has become known as 'climate sensitivity' has stayed remarkably solid ever since: the most recent report of the Intergovernmental Panel on Climate Change pegged the sensitivity as being between 2 and 4.5°C, while adding that higher values could not be excluded.

Although early policy discussions focused on the 550 p.p.m. mark, researchers and politicians soon concluded that such warming would be too much. In 1996, the European Union declared that "global average temperatures should not exceed 2°C above pre-industrial level and that therefore concentration levels lower than 550 p.p.m. CO₂ should guide global limitation and reduction efforts". Over the following decade, 450 p.p.m. became increasingly cited as a level to aim for, because some studies associated that concentration with 2°C of warming.

In their 2008 paper, Hansen and his colleagues offer a number of reasons for arguing that even 450 p.p.m. is too high. The most important are observational: rapid changes in the Arctic and elsewhere have demonstrated that the globe is more sensitive to even today's levels of greenhouse gases than climate models have predicted. Others depend on details of the way climate sensitivity

is defined. The standard approach, going back to Charney's formulation, comes from models that allow fast-reacting components of climate to change but hold constant other, slower factors, such as forests and ice sheets. Yet evidence from the past shows that such slow players are acutely sensitive to varying levels of CO₂ — and are not so slow. By analysing how temperature and greenhouse-gas concentrations actually correlate over the past 500,000 years, as ice sheets have waxed and waned, Hansen and his colleagues find that the true climate sensitivity is 6°C.

Going even further back, the team argues there is evidence for a tipping point in the greenhouse. Some 50 million years ago, CO₂ concentrations were many times today's levels and Antarctica was ice-free. Concentrations declined slowly and crossed a crucial threshold 35 million years ago when the globe was cool enough for an ice sheet to start growing on Antarctica. Through a series of extrapolations, the researchers estimate that the threshold level was between 550 and 350 p.p.m. To avoid any risk of recrossing that threshold and losing Antarctica's ice, best keep at or below the bottom of that range: 350 p.p.m.

Hansen's arguments do not convince everyone. Stefan Rahmstorf of the Potsdam Institute says that there are important distinctions between melting and forming an ice sheet, and the two processes might occur at different greenhouse-gas concentrations. In fact, a 2005

modelling study conducted at Potsdam suggests that during a simulated ice age, the amount of warming needed to melt the North American ice sheet is consistently greater than the amount of cooling needed to grow it². "You have a different threshold for the ice sheets coming and the ice sheets going," says Rahmstorf.

Hansen, though, sticks with the new low figure. He argues that realizing the world is already in dangerous climatic territory "completely changes the story. When you say 450 or 550, you're talking about what rates of growth you are going to allow. When you say we have to get to 350, that means you have to phase down CO₂ emissions in the next few decades."

Peak problems

So how easy would it be to get back to 350 p.p.m.? Most scientists have assumed that it would not take that long to pull down CO₂ levels if humanity went cold turkey and cut off all emissions, says Susan Solomon of the National Oceanic and Atmospheric Administration in Boulder, Colorado. "I've done a little informal poll of colleagues," she says. "It was interesting, the number of smart, knowledgeable people who said if we stop emitting, things will go back maybe in 100 years, 200 tops. But they're not correct. And I didn't believe it would be so long either."

Solomon changed her mind because of a study in which she and her colleagues used what's known as an Earth-system model of intermediate complexity — an EMIC. Although not as detailed as general circulation models, which divide the atmosphere and ocean into millions of cells, EMICs have the advantage of requiring less computing and so can run simulations lasting many centuries. They are also useful because they represent Earth's carbon cycle — the natural movements of carbon between the atmosphere, the biosphere and the oceans. Using an EMIC developed by the University of Berne in Switzerland, Solomon and her colleagues tested what would happen if CO₂ emissions immediately ceased after concentrations peaked at various values, starting with 450 p.p.m. (ref. 3). What they found surprised them. CO₂ levels subsided so slowly that they remained substantially above pre-industrial levels 1,000 years into the future. Global temperatures also stayed up, and had declined only slightly from their peak by the year 3000. In fact the simulations ended before temperatures dropped anywhere close to their starting point.

According to Solomon, the simulated climate recovers so slowly because of

"When you say we have to get to 350, you have to phase down CO₂ emissions in the next few decades."
— James Hansen

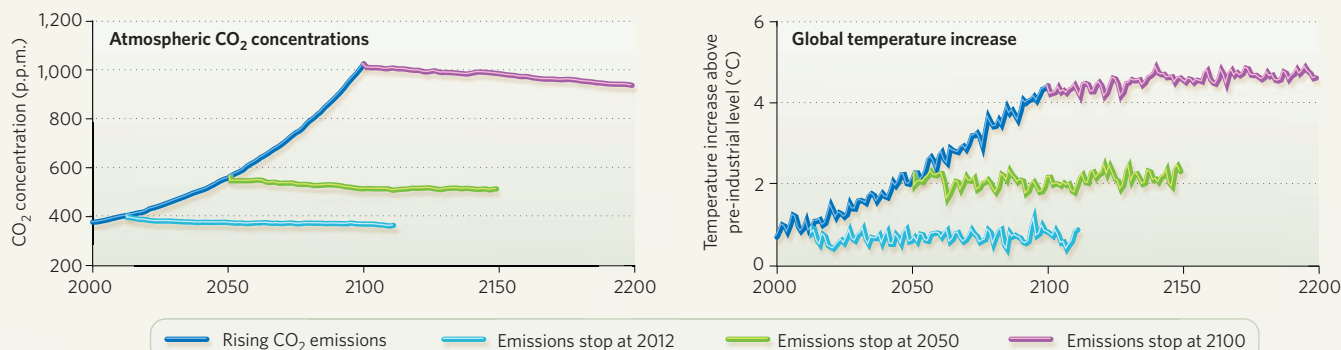


Mark Kibben and Hansen on a 350.org demonstration against coal.

K. DAVISON/GREENPEACE

THE LONG ROAD HOME

A complex computer model shows how Earth might respond if carbon dioxide emissions from humans stopped instantly at various points in the future.



two factors. Natural sinks are only able to take up a fraction of the CO₂ in the atmosphere, so roughly 20% of the emitted gas will stay in the air for at least a millennium, ensuring that it continues to warm the globe long after emissions are cut off. The thermal inertia of the oceans also plays a part: the large mass of ocean water on the planet is delaying the rate of climate warming today because most of it is lagging behind the changes in surface temperature. Once it has warmed it will retard the Earth's cooling after emissions cease.

Slow recovery

Experiments conducted with a more complex model actually make the picture look worse. In a paper this year, Jason Lowe, head of mitigation advice at the UK Met Office, and his colleagues described a study using a general circulation model at the Met Office's Hadley Centre in Exeter, UK, coupled to a carbon-cycle model⁴. He found that after emissions were curtailed, temperatures remained elevated at least to the end of the simulation, which went on 100 years past the cut-off. In fact, if CO₂ concentrations reached 550 p.p.m. or higher before the emissions stopped, temperatures actually increased for at least a century (see graphs above). He would like to see other groups run similar experiments with their own general circulation models.

The take-home message from his and other studies, Lowe says, is this: "If you do end up somewhere you don't want to be, it's probably going to take you a long time to get back to lower temperature levels." Lowe is now looking at how long such warming would last and what kind of trouble it might get the world into by, say, melting the Greenland ice sheet.

Because it will apparently take so long for the climate to recover from excessive warming, researchers are now looking at new strategies to avoid that excess in the first place. One

approach is to stop thinking about the levels at which CO₂ might be stabilized and instead concentrate on something simpler: the sheer amount of CO₂ that can be emitted in total.

In this issue of *Nature*, Meinshausen and his colleagues present results from a coupled climate-carbon cycle model that explores the effects of different emission pathways for CO₂ and the other major greenhouse gases (page 1158). For the period 2000 to 2050, they find that the world would have to limit emissions of all greenhouse gases to the equivalent of 400 gigatonnes of carbon in order to stand a 75% chance of avoiding more than 2°C of warming. Other greenhouse gases, such as methane and nitrous oxide, are expected to produce as much warming as 125 gigatonnes of carbon in the form of CO₂ would; that means emissions of CO₂ itself over the half-century have to add up to less than 275 gigatonnes of carbon. That's an extremely difficult target, admits Meinshausen, considering that emissions over the past nine years have used up almost a third of that allowance already. "Our remaining emission budget is so small," he says. "If we want to have a smooth landing and to decrease emissions in a smooth way, our options are essentially exhausted. We have to bend down our emissions by 2020."

Also in this issue, Myles Allen of the University of Oxford, Meinshausen and their colleagues describe how they ran a series of simulations using a simple combination of climate and carbon-cycle models (page 1163). They find that if humankind could limit all CO₂ emissions from fossil fuels and changes in land use to 1 trillion tons of carbon in total, there would be a good chance that the climate would not warm more than 2°C above its pre-industrial range. Because half of that trillion

tons has already been spewed into the atmosphere, and emissions now average about 9 billion tons a year and rising, the trillion-ton limit would allow the world to follow its current trend for less than 40 more years before giving up carbon emission for good, all at once.

One way of looking at that challenge is put forward by Hansen. Go ahead and burn all the remaining oil and gas in conventional reserves, he says, and at the same time concentrate all efforts on quickly phasing out coal

— or capturing and storing the emissions associated with it. If nations can cut off coal use by 2030 and avoid tapping unconventional fossil fuels, such as tar sands and methane hydrates, the world could limit future CO₂ emissions to 400 gigatonnes of carbon.

Other studies using this total-carbon-emitted approach are now appearing: a couple were presented at the International Congress on Climate Change held in Copenhagen in April. Although differing in details, they come to broadly similar conclusions. Allen says a total limit for carbon emissions, which he calls cumulative warming commitment, is a much more robust figure than a stabilization concentration of CO₂ in the atmosphere.

The problem with looking for a stabilization concentration is that one must first know the globe's long-term response — its 'equilibrium climate sensitivity' — to calculate how much the planet will eventually warm for a given concentration. Estimates of what that equilibrium climate sensitivity might be are shaky, and hence so are forecasts based on it. A focus on total carbon emissions rather than concentrations, however, wipes away that problem because it demands that concentrations go up and — eventually — come back down, never

"If we want to have a smooth landing, our options are essentially exhausted." — Malte Meinshausen

stabilizing at a particular level. So the climate never reaches equilibrium and the uncertainties about its long-term response do not matter as much. "If you assume a finite injection of carbon," says Allen, "you don't need to know the climate sensitivity, so this whole debate about the equilibrium response is moot."

Although the results of the studies might seem too daunting, they do offer a few rays of hope. Andrew Weaver, a modeller at the University of Victoria in British Columbia, Canada, says that in the new studies, what matters is how much pollution goes into the sky, not when it gets emitted. "This allows you some flexibility," he says. From a political perspective, the idea of a cap on total emissions "is a lot easier to get your head around" than a concentration target or, say, a 20% reduction below 1990 emission levels. A cap is like a budget. Once you use it up, there's nothing left to spend.

Unfortunately, the world is behaving as though it expects to be able to arrange a large



Sucking it up

It's simple to mop carbon dioxide out of the air, but it could cost a lot of money. In the second of three features on the carbon challenge, **Nicola Jones** talks with the scientists pursuing this strategy.

When Frank Zeman made a device to mop carbon dioxide out of the air of his laboratory at Columbia University in New York, it didn't look like a machine that could save the planet. Black tape held together plastic parts eaten away by lye; baking soda encrusted the outside. If someone walked behind the air intake (which looked like a grey hair dryer), their exhalations would interfere with the results. But the contraption worked.

Such a device, if scaled up and perfected, could be used to dial back Earth's greenhouse thermostat by taking CO₂ straight out of the sky. Although Zeman's fully functioning desktop device has not yet made it out of the lab, others have developed parts of bigger and more ambitious devices, some of which are heading for commercialization. All are imperfect, but they all work, and that undeniable fact is turning air capture from a 'what-if' pub discussion into a serious proposal.

"Nobody doubts it's technically feasible," says Zeman, now director of the Center for Metropolitan Sustainability at the New York Institute of Technology.

Increasingly it looks like air capture will be needed. Efforts to limit CO₂ emissions will need to be strengthened massively if they are to keep concentrations from reaching dangerous levels, so there may be little choice but to remove some of the CO₂ already in the air (see page 1091) or cool the planet in other ways (see page 1097). "Without having something that is carbon negative, the possibility of avoiding high levels of CO₂ is basically

zero," says Peter Eisenberger, former director of the Lamont-Doherty Earth Observatory at Columbia University and co-founder of the air-capture company Global Thermostat.

In a recent analysis, Roger Pielke of the University of Colorado in Boulder put some numbers on the task ahead. Assuming a middle-range scenario projected by the Intergovernmental Panel on Climate Change (IPCC), humanity must somehow prevent itself from emitting (or must soak up)

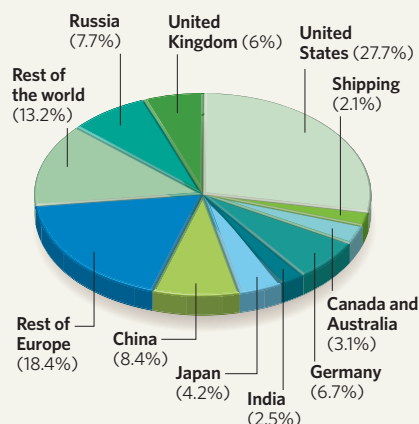
650 gigatonnes of carbon by 2100 to keep concentrations under 450 parts per million (p.p.m.) at that point¹. To put that in perspective, humans added about 9 Gt of carbon to the atmosphere last year.

Economic studies suggest that some reductions could come

affordably, or even at a profit, from fairly obvious places. Deeper cuts would require serious money. A report from the international consultancy McKinsey estimates that energy-efficiency measures, conversion to low-carbon energy sources, and forestry and agriculture management could — with serious effort — cut about 10 Gt of carbon emissions annually by 2030, for under US\$300 per tonne. But it will be much harder and more expensive to get at any fraction of the remaining 9 Gt of annual emissions expected that year in a business-as-usual scenario². Pielke is one of many beginning to wonder whether mopping up CO₂ with chemicals and machinery — a strategy with an ironically un-green image — might be part of the answer.

It could be an unbeatable idea. Sponging CO₂ from the air has a direct, immediate and measurable effect on the source of the

CUMULATIVE CO₂ EMISSIONS 1750–2006



overdraft. And researchers can only come up with so many ways of presenting the gravity of the carbon problem to the rest of the world. "At some point, you begin to throw your hands up. It's very frustrating," says Weaver, who pulls a reference from an ancient global crisis. "Climate scientists," he says, "have begun to feel like a bunch of Noahs — thousands of Noahs." ■

Richard Monastersky is a features editor with *Nature* in Washington DC.

See also Editorial, page 1077, and www.nature.com/climatecrunch.

1. Hansen, J. et al. *Open Atmos. Sci. J.* **2**, 217–231 (2008).
2. Calov, R. & Ganopolski, A. *Geophys. Res. Lett.* **32**, L21717 (2005).
3. Solomon, S., Plattner, G.-K., Knutti, R. & Friedlingstein, P. *Proc. Natl. Acad. Sci. USA* **106**, 1704–1709 (2009).
4. Lowe, J. et al. *Environ. Res. Lett.* **4**, 014012 (2009).



problem, avoiding the possible side effects of geoengineering. Air-capture devices can be sited anywhere, although preferably on cheap land with an untapped renewable energy supply and a geological reservoir that could serve as a dump for the captured gas. In principle, there is no limit to how much CO₂ you can extract: name an atmospheric concentration you'd like to end up with, and the technology can get you there.

To many in the 1990s, that cost seemed ridiculously high. In engineering circles, the dogma ran that the ease of extracting a gas was proportional to its concentration. At 0.04%, CO₂ in the atmosphere seemed exceedingly difficult; the effort and money needed to extract and store CO₂ from industrial flue stacks, where it can make up perhaps 10% of all gas, is already high, estimated by the IPCC to cost between \$70 and \$260 per tonne of carbon (see *Nature* 442, 620–623; 2006). The assumption was that filtering CO₂ out of the atmosphere would be 250 times harder and vastly more expensive.

That assumption turns out to be wrong. The benefit of air capture is that it deals with a nearly infinite and relatively clean source, so there is no need to scrub out polluting gases

before beginning and no need to take out every last bit of CO₂. Thermodynamically, the task proves to be about twice as hard as flue-stream capture³. Better still, the technology to make such devices is already available.

Although air capture has been ignored by the IPCC and sidelined by scientists, that is changing. Researchers in Canada, the United States and Switzerland have come up with plans, tested prototypes, filed patents and founded companies to pursue the idea.

Which technology will win out is yet to be seen. The Virgin Earth Challenge, launched by airline entrepreneur Richard Branson and former US vice-president Al Gore in February 2007, offers up to \$25 million for the first demonstrably viable commercial design to remove significant amounts of greenhouse gases from the atmosphere (the exact criteria are unclear). As yet the prize goes unclaimed.

The bare-bones chemistry of carbon capture is simple. The simplest thing to do is to expose air to a sorbent of lye (NaOH). This reacts with CO₂ to create a solution of sodium carbonate. It's so simple that Klaus Lackner, also of Columbia University, once helped his daughter to do it for a school science project. To get the

carbon out of solution, a trick can be borrowed from the pulp and paper industry: when slaked lime (Ca(OH)₂) is added to the mix, particles of calcium carbonate settle out. Throw this into a kiln and you are rewarded with a pure stream of captured CO₂ and quicklime (CaO), from which the sorbent can be renewed.

Crude prototypes

This is how Zeman's desktop device worked, and also how David Keith of the University of Calgary in Alberta, Canada, is pursuing the problem. Keith built a large-scale machine a few years ago to see how much CO₂ could be sucked up in practice. He calls it the 'Russian tractor' technique — not especially high-tech, but proven to work. A prototype featured on the Discovery Channel in 2008 mopped up a few kilograms of carbon overnight.

Keith didn't build the second half of the scheme — the 900 °C kiln that spits out concentrated CO₂ — because that's already a known industrial process. It's also the energy-intensive and costly part. Nevertheless, he is setting up a company called Carbon Engineering, convinced the idea is worth pursuing, and is working to reduce costs.

Keith has chosen the most obvious approach to the problem but admits that others have "much more clever" schemes. That includes a material being developed by Lackner for the company Global Research Technologies, based in Tucson, Arizona, and funded by a \$5-million donation from the late billionaire Gary Comer. (Comer, founder of the Lands' End clothing-catalogue company, donated money to fight climate change after he sailed through the Northwest Passage in 2001 without being blocked by ice.) In April 2007, Global Research Technologies had its first demonstration of air capture with a prototype device. It was a success, widely lauded in the press, but it needed further work. For one thing, it just vented the captured carbon out the back. For another, it didn't behave as it was expected to. "When we closed the door on it, something was happening we didn't understand," says Lackner.

The device used a commercially available

A way to pay for capturing carbon dioxide

Around the world, some 5.5 million tonnes of carbon are used each year in dry ice or in compressed carbon dioxide to transport ice cream, flash-freeze meat, blast-clean engine blocks and carbonate drinks. The CO₂ purchased for that 'merchant market' can cost from US\$130 to \$1,100 per tonne of carbon.

Capturing CO₂ from the air is estimated to cost up to \$500 per tonne of carbon using today's technologies. If the aim was to bury the carbon underground and sell the deficit on the carbon-credit

market, it would be hard to make a buck: carbon prices are currently €25 (US\$30) per tonne of carbon on the European trading market. But the \$1,100 per tonne price tag within the merchant market sector looks appealing to Klaus Lackner and the Global Research Technologies air-capture company.

Lackner foresees a world where the company's not-yet-built air-capture devices are carted around the United States from one willing customer to another. Because the CO₂ used by the merchant

market ends up back in the air, this wouldn't do much to help the planet. But such a market could drive technological development and lower the price of air-capture devices, Lackner argues, until it becomes profitable to fight climate change.

An alternative profit-making scheme would be to turn captured CO₂ back into hydrocarbon fuel. Again, that's not ideal for reducing atmospheric CO₂, but it does create a carbon-neutral way of keeping fuel-guzzling cars on the road.

N.J.



STONEHAVEN PRODUCTIONS

Klaus Lackner has imagined huge 'farms' featuring thousands of air-scrubbing devices that could soak up billions of tonnes of carbon from the atmosphere.

wet resin to mop up CO_2 . When its designers analysed the results, however, they realized the material was better than they thought. Not only did it turn CO_2 into carbonate, but in a dry environment it would go a step further to bicarbonate. When they exposed the resin to water, the bicarbonate flipped back to carbonate, releasing CO_2 and water vapour. They didn't need a kiln — they just needed to expose their loaded resin to water in a relative vacuum, and then pressurize the result to condense the water out. "All you pay for is making the vacuum, pumping and pressurizing," says Lackner.

Others argue that kiln-temperature heat isn't necessarily a problem. In Zurich, Aldo Steinfeld and colleagues at the Swiss Federal Institute of Technology are using the Sun-tracking mirrors used by solar-power plants to heat up their air-capture reactor to 800°C . They have a fully functioning lab model, and hope to have a larger field prototype within a few years to hand to an industrial partner.

Eisenberger, on the other hand, needs only low temperatures — under 100°C , achievable using waste heat from power plants or cement factories — to run his system. Eisenberger's company Global Thermostat, which was founded in 2006 with Graciela Chichilnisky, an economist at Columbia University, is waiting for venture-capital funding to make a prototype,

which could come as early as this autumn.

Eisenberger imagines a future in which air-capture devices start to be deployed by 2015; by 2020, half of new power generators are matched with air capture, and by 2040, some 9 Gt of carbon is being pulled from the air per year, to a total of 650 Gt by 2100 — the amount that Pielke also estimated would be needed. (Coincidentally, that total roughly matches the IPCC's estimate of the Earth's geological capacity to act as a garbage dump for buried gas). This whole operation could be accomplished by, say, 35,000 facilities that each took a quarter of a million tonnes of carbon per year out of the air. The combined footprint of this global operation would total less than 300 square kilometres — a

fraction of the size of London.

Because Eisenberger assumes the world will also make substantial cuts in emissions over the same period, his air-capture scenario would return atmospheric concentrations to 380 p.p.m. of CO_2 by 2100, and they would continue to decline thereafter. The price? About \$60 trillion for the air capture, or roughly \$660 billion per year. That's on the same scale as the US economic stimulus package against the current recession, but every year for a century.

The price is the hardest thing to estimate, since no one has yet built a full-scale device. When Lackner first put out figures of about \$100 per tonne of carbon in 2006, many saw

it as massively over-optimistic — some joked that the real price was one mysterious 'Lackner' per tonne, given the apparently magical capacities of his material, the identity of which was kept under wraps for commercial reasons at the time. Today, Eisenberger's estimate is slightly cheaper still.

Cost competitive

At the other end of the scale, Keith has estimated it might cost \$500 per tonne of carbon using today's technologies. That would rack up a bill of \$325 trillion to soak up 650 Gt of carbon, but Pielke notes that such a price tag would still only be 2.7% of global economic output by 2100. That compares favourably with price estimates of the IPCC (–1 to 5% of global economic output) and economist Nicholas Stern (–2 to 4%) for stabilizing air concentrations at 450 p.p.m. without air capture.

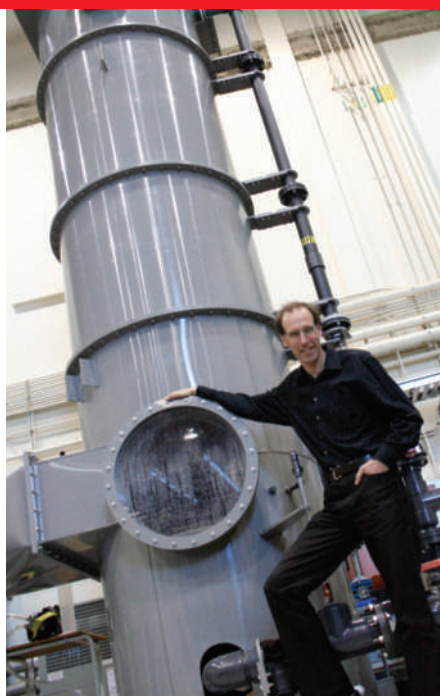
"We should be looking into it, at least," Pielke concludes. To put the cost issue in perspective, he notes, if all the emissions from US cars were sucked up by air capture using today's technology, and the cost tagged onto the price of petrol, motorists in the United States would still have one of the lowest pump prices in the world.

Many air-capture enthusiasts talk about countering something on the scale of global aircraft emissions, projected to reach about 0.25 Gt of carbon per year by 2030. (Technology can reduce carbon emissions from power plants and cars, but it is difficult to reduce such emissions from planes.) This is where

"It is the most expensive climate-mitigation technology. And that's a good thing."
— Frank Zeman

Roger Aines of Lawrence Livermore National Laboratory in California sees air capture playing a potential part. He and his colleagues are making an overview assessment of the strategy, and estimate that the quarter-gigatonne target could be met by, say, a thousand 250,000-tonne air-capture facilities requiring a total of 900,000 gigawatt-hours of energy per year. This is slightly more than the total electricity generated by the 104 nuclear power plants in the United States. If wind were to supply the power, the world would need something like 135,000 additional 1.5-megawatt turbines. That would approximately double the current global wind-power capacity.

Such a scenario is within the realm of possibility, but it demands an increase in energy production just at a time when we should be trying to break our energy addiction. For some, that's a critical problem. Every dollar spent on air capture instead of shifting to renewables is "a long-term loss to society", says Mark Jacobson of Stanford University in California. His concern is that researching a 'get out of jail free' card for climate change would provide an excuse to continue unabated emissions.



David Keith and his carbon-capture machine.

That worry is voiced by many, but it is also dismissed by many. "For some people there's concern that if there's hope that air capture will work, it reduces the incentive to reduce emissions," says Pielke. "That makes as much sense as saying we shouldn't have open-heart surgery because it stops people from

lowering their cholesterol. We need both."

No one argues that air capture is a cure-all. Eisenberger sees it as a necessary bridge to get us more painlessly to our goal of a renewable-energy economy. Despite the 'reasonable' price tag of air capture, it is still cheaper, and more sensible, to capture large-industry pollutants at source and to reduce energy use. "Air capture would be a back-stop technology to fill in the gap between what we can achieve and what our goals are," says Pielke.

"It is the most expensive climate-mitigation technology," agrees Zeman. "And that's a good thing. It has this role as the upper bound on solving the climate problem." No matter what we have to do to get the atmosphere settled, it won't cost more than this.

Nicola Jones is a commissioning editor for *Nature's Opinion* section.

See also Editorial, page 1077, and www.nature.com/climatecrunch.

1. Pielke, R. A. Jr *Environmental Science & Policy* (in the press).
2. McKinsey & Company *Pathways to a Low-Carbon Economy: Version 2 of the Global Greenhouse Gas Abatement Cost Curve* (2009).
3. Keith, D. W., Ha-Duong, M. & Stolaroff, J. K. *Climatic Change* **74**, 17–45 (2005).

K. BENDIKTSEN, UNIV. CALGARY

Great white hope

Geoengineering schemes, such as brightening clouds, are being talked about ever more widely. In the third of three features, **Oliver Morton** looks at how likely they are to work.

Something utterly insubstantial is rising above the rim of the beaker on the table. It looks like a white mist; it feels like nothing. Run your hand through it and you get no sense of warmth or cold. It leaves no moisture on the skin, no smell, no taste. It's just a whiteness.

You can see that it would spur curiosity; that it might spur controversy is harder to imagine.

The mist is made up of droplets of water just a few micrometres across, thousands of times smaller than a raindrop. The man who set up this beaker as a demonstration, a nominally retired professor of engineering at the University of Edinburgh, UK, named Stephen Salter, thinks that ships designed to produce such mists could whiten the low layers of cloud that hang above the sea over large areas of the globe. Established theory predicts that such whitening, if achieved, could cool Earth significantly — a thousand such ships might cool it as much as decades of carbon dioxide emissions would warm it.

The beaker demonstration was part of a one-day meeting held at the University of



Edinburgh in mid-March to look at how cloud whitening could move beyond the era of the tabletop. The meeting's agenda was vast, encompassing climate modelling, cloud physics, data from a field campaign studying clouds off the coast of Chile, the design of ships and the min-

utiae of the tiny nozzles needed to create such ultra-fine sprays. It ended up, as almost all such discussions of cooling the Earth do, heading off into questions of morality, politics and public perception.

The frequency of such meetings shows how this topic, known as geoengineering, is gaining, if not acceptance, at least an enhanced currency. For a number of the participants, this was their second day of geoengineering presentations that week — there had been an all-day discussion of the topic at the International Scientific

Congress on Climate Change in Copenhagen two days before. The following week, some of the key players would be at it again, this time at a workshop organized by the US Defense Advanced Research Projects Agency in Stanford, California.

As yet, though, these discussions are, like Salter's mists, insubstantial. Very little funding is available for real research into whether

ships are the best way to whiten clouds, or whether cloud whitening is really a workable way to cool the world. And that is cause for concern because there is a real possibility that such schemes won't work. "The most dangerous case is ... when you think that geoengineering works and you're wrong," said David Keith of the University of Calgary in Canada while at the Copenhagen meeting.

The worry that Keith and others share is that a growing interest in geoengineering



Could a fine mist help to combat global warming?

R. GUSTHART

tends to promote the belief that it is a plausible option, even though there is not enough research to back that up. And interest is undoubtedly growing, in ways that go well beyond the small number of scientists trundling from meeting to meeting. The American Meteorological Society in Washington DC is consulting with its membership about a policy statement on the subject, and Britain's Royal Society is preparing a report chaired by John Shepherd, an oceanographer at the University of Southampton, UK. John Holdren, science adviser to US President Barack Obama, says that the technology should be looked at on the basis that nothing should be taken off the table; the economist Nicholas Stern, author of the influential *Stern Review on the Economics of Climate Change*, says much the same in an interview on this week's *Nature* podcast.

The discussions in science and policy circles are typically and appropriately couched with caveats about the unknown feasibility and safety of such ideas and the much greater desirability of cutting emissions. Although Greenpeace International has not called for a ban on geo-engineering research, David Santillo, a scientist with the organization, argued at the Copenhagen meeting that such a policy might ideally be the best position, so that no creeping faith in the possibility of a last-ditch alternative would ever undercut the need to reduce emissions.

First, choose your wavelength

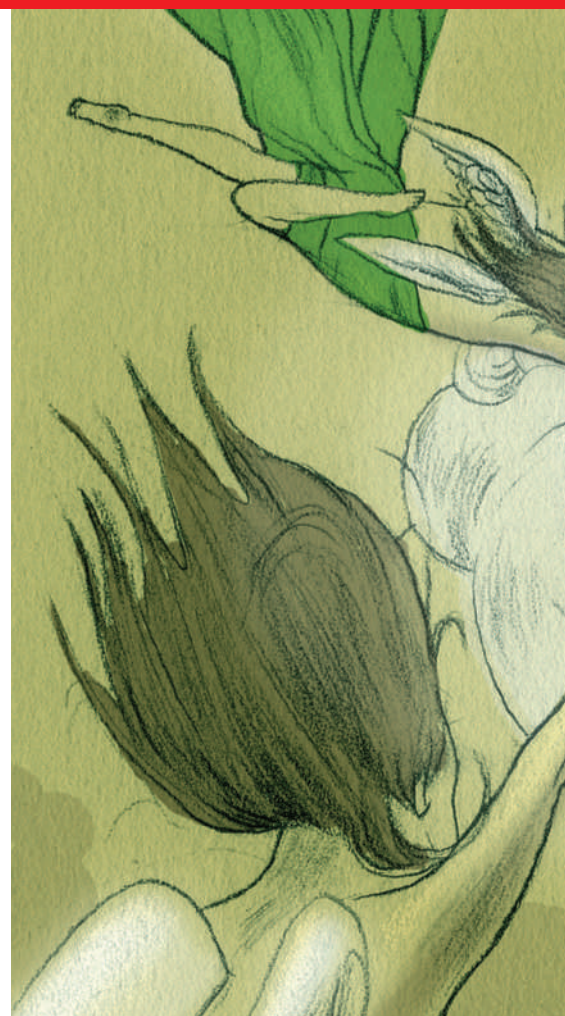
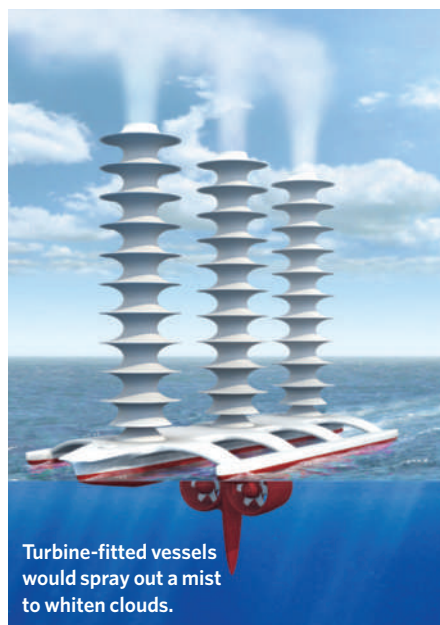
Geoengineering approaches can be divided into two categories, short-wave and long-wave. Short-wave approaches reduce the amount of energy entering Earth's system by increasing the amount of sunlight that bounces back out into space without being absorbed. Long-wave approaches help infrared radiation to escape from the atmosphere, usually by reducing the concentration of CO₂ — for example by turning biomass into charcoal and burying it, or by fertilizing plankton blooms. Both approaches have their proponents, but as Timothy Lenton and Naomi Vaughan of the University of East Anglia, in Norwich, UK, have shown, the amount of cooling that might be expected from long-wave schemes is substantially less than you can get from short-wave ones (T. M. Lenton and N. E. Vaughan *Atmos. Chem. Phys.* **9**, 2559–2608; 2009).

Short-wave schemes can be distinguished further by altitude. Sunlight can be reflected before it reaches the planet by some sort of shield in space. Or it can be turned away closer to home: by aerosol particles in the stratosphere; by clouds in the lower atmosphere; or by white objects on Earth's surface, such as painted buildings and roads. Lenton and Vaughan found that the last of those options — making the surface

more reflective — would not be able to produce an effect large enough to counteract a doubling of CO₂. The first — 'sunshades in space' — could in principle counteract any warming effect that can be imagined. But it would require spacecraft vast in either size or number, a spectacularly ambitious and costly undertaking for a civilization that has no way of delivering 100-tonne payloads to the lowest of orbits.

That leaves the stratosphere and the clouds. The stratosphere has received the bulk of geo-engineering attention (see *Nature* **447**, 132–136; 2007), mainly because volcanic aerosols at that height have clearly cooled the globe, and because this type of cooling is amenable to the sort of global modelling that climate researchers are good at. But such an intervention raises numerous issues. Particles injected into the stratosphere might also catalyse chemical reactions that could deplete the ozone layer. And although they — or the raw materials to produce them — could be lifted that high in a number of ways, it could be hard to make such a system operational and potentially impossible to ensure that the particles remain at the right size. Furthermore, stratospheric aerosol particles have a lifetime of a couple of years — long enough to wreck whole growing seasons if their side effects included suppressing rainfall.

An attraction of the cloud-based approach is that it gets around some of those issues. It uses nothing more than seawater, it doesn't require things to be lifted tens of kilometres into the sky, it can be tested on a local or regional scale and it can be turned off instantaneously. The idea was first put forth in 1990, by John Latham, a British atmospheric scientist now based at the National Center for Atmospheric



Research (NCAR) in Boulder, Colorado. For water vapour to form into clouds, the atmosphere needs to contain particles for the water vapour to condense on, called cloud condensation nuclei. Air that is well supplied with these nuclei will contain many small water droplets. In air that has a smaller number of nuclei, condensation will form fewer, larger drops. Clouds consisting of small droplets are more reflective than those with larger ones, and under some conditions they will also last longer. So if you were to inject condensation nuclei into clouds you would, other things being equal, make them brighter.

Some 25% of the world's oceans are covered with thin, low-lying layers of stratocumulus cloud. Make them brighter, Latham says, and you could cool the planet. So Salter has designed wind-powered ships that use underwater turbines to drive machinery to make the fine mist that can provide the necessary particles. Such ships, which would operate, in principle, without a crew, could seed clouds over the necessary swathes of ocean.

Phil Rasch, a climate modeller who until recently was a colleague of Latham's at NCAR and who is now at Pacific Northwest National Laboratory in Richland, Washington, has looked at how effective such brightening might be. By manipulating the number of cloud condensation nuclei in a global circulation model of the climate, Rasch found that seeding 25–50%

ILLUSTRATIONS BY J. BURTON



of the ocean with droplets of the right size could offset a greenhouse warming of 3 watts per square metre — the amount that might be expected from a doubling of CO_2 . Running the model with doubled CO_2 showed that seeding could keep the average temperature where it is now, or in some scenarios actually diminish it. This might take hundreds or indeed thousands of ships, depending on their capacities. Salter, for one, thinks that such a fleet would be quite cheap. “Do not be put off by yachts for the rich, with diamond-encrusted lavatory seats. Check out fishing boats.” He thinks that around a thousand \$2-million ships could do the trick.

But even though they are of the same order of magnitude, the cloud effect does not perfectly counter the greenhouse effect. Brighter clouds cool only during the day — and do it best in summer — whereas greenhouse warming is felt 24/7. This imbalance applies to short-wave geoengineering schemes that use stratospheric aerosols, too, and it means that the net effect seen in models that include both greenhouse warming and geoengineered cooling of this type would never be just the status quo ante. Crucially, patterns of precipitation, among other things, change. But the disparity would probably be much stronger in a cloud-brightening scheme that targets only the ocean than in a stratospheric veil that operates globally. A wide range of climate phenomena are driven by temperature differences between

the oceans and the land, from sea breezes to monsoons. How they would be affected by cloud brightening is something no one can yet say with any confidence. Nor can they say what the implications might be for ocean currents — a topic that Rasch is actively pursuing.

Clouding the issue

Atmospheric scientists understand fairly well the process that creates and dissipates the clouds in question. But the relative weights of various processes and how they interact are still not understood — and those weights might be crucial to whether a cloud-seeding scheme could actually work, says Tom Choularton, who heads the cloud-research group that Latham started at the University of Manchester, UK, in the 1960s. Take, for instance, the role of convection in the marine boundary layer — the mixed layer of air that extends a few kilometres above the ocean surface. At night, convection mixes air from the bottom of the layer all the way to the top, lifting up moisture from the warm ocean to form clouds at the top of the layer. During the day, though, the clouds at the top of the layer bask in the sun, so the layer is heated from both the top and the bottom. This means that the whole-layer convection pattern can break down, with the upper part, including

the clouds, ‘decoupling’ from the lower. The evaporation of drizzle beneath the cloud can cool the middle of the layer, exacerbating the effect. Decoupling would make it much harder to get condensation nuclei from the surface to the clouds in the first place.

A linked concern, pointed out by Rob Wood of the University of Washington in Seattle, is that it is hard to gauge the effect of adding condensation nuclei on the clouds’ lifetime. More cloud nuclei would be expected to make clouds persist by suppressing the growth of large droplets, which fall out of the cloud as precipitation. But suppressing precipitation might increase

the vigour of the circulation in the upper part of the boundary layer. That could draw in dry air from above, leading to increased evaporation, which would thin and disrupt the cloud.

The approach that Latham and Salter have laid out relies

on the creation of very small water droplets, which dry out to form little specks of sea salt on which the water vapour in the cloud will condense. These salt particles will attract more water vapour than the particles that currently dominate the aerosols over the open ocean, which tend to be ammonium sulphate created by chemical reactions in the atmosphere. If the salt particles outcompete the sulphate particles, the net effect on droplet number might be small

“Do not be put off by yachts for the rich. Check out fishing boats.”

— Stephen Salter



Whitening the clouds that lie low over the ocean could help to cool the Earth.

or even negative, with bigger droplets forming.

To avoid this fate, the salt particles need to start off very small indeed. That makes designing the system to create the spray even more challenging. Salter has been told by various experts in industrial spraying that it would be simply impossible. He refuses to believe that, and the Edinburgh meeting saw vigorous debate between him and Armand Neukermans, a California-based innovator with a long track record of developing technology for, among other things, ink-jet printers, about ways to make very small droplets. They agreed that the smallest size feasible remains unclear — as does a great deal else associated with making such droplets day in day out, starting with unfiltered sea water, on a ship with no human crew.

VOCALS support

At every point in the Edinburgh meeting, and at every scale from the micrometre to the global, the need for further research shone through. Many of the participants had been involved in a project that might serve as a model for future studies. In late 2008, researchers from 30 institutions used a range of satellites, aircraft, research ships and land-based observations to study cloud processes off the coasts of Peru and Chile, as part of a project called VOCALS, which is a component of an even larger study called the Variability of the American Monsoon Systems (VAMOS) project. (VOCALS, on which Wood was the principal investigator, stands for the VAMOS Ocean-Cloud-Atmosphere-Land Study). The data they gathered should improve understanding of the interaction between clouds, drizzle and aerosols, and the degree to which

aerosols both above and below the clouds affect their properties. Of particular interest were holes that form in the cloud layer as a result of local decoupling.

Daniel Rosenfeld of the Hebrew University in Jerusalem sees opportunity in such holes. He thinks that the aerosol concentrations will differ very little between a solid cloud bank and one filled with holes. In a poster at last December's meeting of the American Geophysical Union in San Francisco, he built on this idea to suggest that aerosol-producing ships might be able to convert a patchwork cloud layer

into solid cloud quite easily. This already seems to happen sometimes in the smoky wakes of commercial ships. If it works, such a strategy would provide a much more powerful cooling effect than merely brightening clouds that are already there, giving Rosenfeld's approach remarkable leverage, he thinks, with a much smaller fleet needed. But as the more intense cooling has a more local effect, the equal-not-opposite mismatch between cooling and warming might be even more problematic.

For pretty much everyone at the Edinburgh meeting, the medium-term goal for research into cloud brightening seemed to be a VOCALS-type experiment in which droplet-making technologies and their effects are studied over hundreds of square kilometres. Even that would not resolve all the issues about the feasibility of such a scheme, and it can't be rushed into — more basic research needs to be done first, in terms of cloud modelling and nozzle making — but it would be a start. And

such tests could have implications for climate research more broadly. As Keith pointed out at the meeting, "There's the potential to learn a lot more by intervening." Being able to experiment on clouds might reveal a great deal that climate scientists need to know, about what will happen in a warmer world even if geoengineering proves impossible, Choularton says.

However, as yet there is no funding for such efforts. Of everyone using the VOCALS data, only one PhD student is doing so with a specific geoengineering-oriented goal. And what concerns researchers is that things might stay that way. It is one thing to get people to talk about geoengineering — it is another to make it a serious research topic with significant funding. Indeed, if general discussion of the possibility leads to increased polarization and opposition — as it very well might — then it may become more difficult, not less, to do the work necessary to test the possibility.

Paradoxically, the intense debate over this topic could keep it alive and in the realm of possibility. At some point in the future, perhaps not too far away, society could be searching for a last-gasp response to global warming. The more that people talk about geoengineering, the more likely they will be to assume there is something solid in the idea. In fact, though, it may be as insubstantial as Salter's mist aspires to be. ■

Oliver Morton is Nature's chief News and Features Editor.

Since the writing of this article, the author now participates in scientific research on the topic.

See also Editorial, page 1077, and www.nature.com/climatecrunch.

"There's the potential to learn a lot more by intervening."
—David Keith



CORRESPONDENCE

Stem-cell treatments for spinal-cord injury may be worth the risk

SIR — In his Correspondence 'Caution urged in trial of stem cells to treat spinal-cord injury' (*Nature* **458**, 29; 2009), Yves Barde questions the wisdom of testing oligodendrocyte precursors derived from embryonic stem (ES) cells in patients, despite the promise that such cells hold for repairing these injuries in rodents.

After traumatic injury to the spinal cord, the axons adjacent to the lesion often remain intact but become demyelinated (J. Silver and J. H. Miller *Nature Rev. Neurosci.* **5**, 146–156; 2004). Stem cells derived from adult or embryonic sources can remyelinate denuded axons and restore limited, but significant, recovery of function (see, for example, M. Sasaki *et al. Prog. Brain Res.* **161**, 419–433; 2007). It is therefore plausible that treatment with oligodendrocyte precursors might perform similarly in humans.

I concede Barde's point that there are other issues to consider in developing a comprehensive treatment for paralysis, including axon regeneration, axon-growth inhibitors and the lesion scar. But even partial restoration of function through remyelination would yield valuable improvements in patients, including an ability to breathe independently, enhancement of hand dexterity, recovery of sexual function and restoration of bowel and bladder control.

There is evidence for some degree of recovery from spinal-cord injury in animals after experimental cell transplantation alone, or in combination with other agents. But as the rest of the world moves forwards in developing human treatments in this area, the United States lags behind. Given the severity of such injuries and the often-overlooked fact that a patient's condition deteriorates steadily over time, a calculated degree of risk may

be justified in pioneering clinical efforts to resolve the dilemma of lifelong paralysis.

People from the United States with spinal-cord injuries are flocking to clinics around the globe for cell transplantations — despite warnings from the scientific and medical communities about the potential dangers — largely because of the paucity of treatment options at home. The US Food and Drug Administration's forthcoming safety trial using ES cells in humans with spinal-cord injuries (see *Nature* **457**, 516; 2009) is therefore a long-awaited and welcome first step.

**Jesse Owens Biomedical Program,
University of Alaska, Anchorage,
Alaska 99508, USA**
e-mail: afjlo@uaa.alaska.edu

A lesson or two from a regional economic argument

SIR — In his Commentary on how to survive the recession, 'Work for the greater good' (*Nature* **457**, 959–960; 2009), Eric Rauchway discusses the role that science and technology had in improving living conditions in the Tennessee Valley in the 1930s. As a former staff historian for the Tennessee Valley Authority (TVA), I would like to highlight factors debated then that could still be pertinent today.

As Rauchway tells us, the innovative agricultural reform measures led by Harcourt Morgan helped to restore economic and environmental health to a segment of the country that had suffered for decades, long before the Depression. But there was controversy over the place of science and technology within the TVA's reform agenda, particularly between the other two TVA board members: David Lilienthal, a lawyer who had served on the Wisconsin public service commission, and Arthur Morgan (no relation to Harcourt), an engineer who had been president of Antioch College in Yellow Springs, Ohio.

For Lilienthal, the TVA's mandates for flood control and resource conservation were secondary to one thing: power generation. Only a small percentage of the valley's rural population had access to electric power, and Lilienthal wanted the TVA to produce cheap and reliable electricity to put an end to poverty in the region. Arthur Morgan had even bigger plans: he wanted science to take a back seat and the TVA to be a regional planning agency, with a focus on social and economic reforms, including alternative land-use schemes to encourage the growth of small villages and ensure long-term conservation of resources.

This conflict undermined the TVA's ability to function effectively and, in 1938, President Franklin Roosevelt dismissed Arthur Morgan. The president's decision sealed the agency's fate as an institution primarily focused on electric-power production and distribution. Today, the agency's other responsibilities are dwarfed by the management of its US\$9-billion power system.

But 2009 is not like 1929 (at least, not yet), President Barack Obama's recovery plan is not the New Deal, and no broadband valley authority is destined to do for high-speed Internet connections what the TVA did for access to electricity. However, there are parallels between the debates that surrounded the TVA and those now taking place in Washington DC. What role should the public sector have in the nation's economic-recovery efforts? Should government policies be directed towards getting the economy back on its feet, towards radically changing its direction, or both? Do solutions lie in politically free applications of science and technology? Or should a new social vision determine how — and to what extent — science and technology are applied to realize these goals?

The bankers and modellers who thought that they could vanquish history through the use of the latest financial instruments

have failed us miserably. The past has returned with a vengeance. Experiences such as those of the TVA can teach us a thing or two about how to move ahead responsibly in the face of today's global financial crisis.

Daniel Schaffer Sesto al Reghena, Italy
e-mail: ddschaffer@yahoo.com

Romanian funding cuts call for more stringent criteria

SIR — The Lisbon summit in 2000 persuaded many governments that it was in the interest of Europe's long-term economic growth to restore priorities in science and research expenditure; this would also help found the next generation of researchers, innovators and technicians. But these laudable aims are being undermined by the current economic crisis, which disproportionately affects the Eastern European economies.

Despite generous support for all branches of science in the two years leading up to Romania's accession to the European Union in 2007, the latest financial situation is impeding progress just as it was starting to gather momentum. The research budget for this year has been cut back to about 40% of what it was in 2008. Successful projects that won funding last year are slowing to a halt, along with the ascent of academia and industry.

But it is not just a matter of curbed expenditure. Science lobbyists and policy-makers should have used the financial crisis as an opportunity to apply more stringent funding criteria to raise the quality of scientific output and accelerate progress. Romania cannot actively compete in today's scientific and economic arenas without stimulating scientific creativity.

Tudor Luchian Laboratory of Biophysics and Medical Physics, Department of Physics, Alexandru I. Cuza University, Iasi, 700506, Romania
e-mail: luchian@uaic.ro

COMMENTARY

Overshoot, adapt and recover

We will probably overshoot our current climate targets, so policies of adaptation and recovery need much more attention, say **Martin Parry, Jason Lowe and Clair Hanson.**

If policy-makers are to reach international agreement on greenhouse-gas emissions at the United Nations Framework Convention on Climate Change conference in Copenhagen in December, they need to be optimistic that their decisions could have swift and overwhelmingly positive effects on climate change. The reality is less certain, but no less urgent.

Even the most restrictive emissions policies proposed to date leave a sizeable chance that significant climate change will occur over the next several decades, probably surpassing the 2°C warming target adopted by the European Union and held by many as a dangerous limit beyond which we should not pass¹. We must therefore complement a strong emissions policy with a plan to adapt to major environmental, social and economic changes in the lengthy period during which we will overshoot safe levels of climate change. This will require much more investment in adaptation than is currently planned.

The stringent greenhouse-gas-reduction policies posed at the G8 summit in 2007, for example, assume a reduction of around 50% in emissions by 2050. The storyline we develop here assumes that everyone agrees to this target at the Copenhagen talks in December and that policy is implemented immediately, thus ensuring the start of a downturn in global emissions — currently increasing at about 3% per year — by 2015 (Fig. 1). Implementing the



policy would mean continual 3% year-on-year emissions reductions that could, after several centuries, lead to greenhouse gas concentration of about 350 parts per million (p.p.m.) of carbon dioxide equivalents. A new and useful approach for quantifying long-term emission targets is presented in two new pieces of work published in this issue (pages 1158 and 1163).

We have simulated the outcomes of this 3%-per-year reduction strategy with a simple Earth system model² and have plotted them on a table of projected effects that we constructed, with other Intergovernmental Panel on Climate Change Working Group II authors, for the IPCC 2007 assessment³ (Fig. 2). Our storyline — immediate implementation, achieving peak emissions in 2015 and 3% global emissions cuts annually thereafter — leaves an even chance of exceeding 2°C of warming. Temperatures would probably peak around 2065 just above a 2°C rise, but with about a 20% chance of exceeding a 2.5°C rise. If the same rate of year-on-year emissions reductions was maintained over the next century, temperatures would slowly recover to about 1°C of warming by 2300. This would be a considerable challenge, however, because it would require substantial reductions in fossil fuel use and deforestation and, in the long term, major and much more difficult reductions of emissions from agriculture.

With the same 3%-per-year long-term emissions reductions but a slower start, peak temperatures would rise substantially and the overshoot would extend. For example, delaying mitigative action by ten years and so reversing emissions trends by 2025 would raise peak median temperature by about 2.5°C; delaying by a further ten years (a 2035 downturn) would mean a rise of about 3°C, with much longer recovery.

"We should be planning to adapt to at least 4°C of warming."

The damage from these levels of warming could be substantial, placing billions more people at risk of water shortage and millions more at risk of coastal flooding. To avoid such damage will require massive investment in adaptation, such as improv-

ing water supply and storage, and protecting low-lying settlements from rising seas. But how much adaptation should we plan for?

It will be very expensive to protect against warming at the upper end of the uncertainty range. We therefore will need to make a judgement about what damage is worth avoiding completely and what we will have to bear. Looking at the median projected warming for different peak emissions dates, with ranges of uncertainty of climate response (shown by horizontal bars in Fig. 2), one can predict that damages to the right of the median should probably be avoided by mitigation, while those to the left would probably need to be borne or be adapted to.

If we make some simple assumptions about the amount of risk we wish to cover, we can identify how much we need to adapt. For example, we might assume that small amounts of adaptation would cover at least 10% of the risk of harm; moderate amounts might cover half; and much larger amounts could cover 90%.

The timing and stringency of emissions reduction will also influence the scale of potential damage, which would affect how much adaptation is needed: slower and lower reductions would lead to larger effects. Thus, if we wished to adapt to 90% of the risk implied by delaying mitigative action until 2035, we should be planning to adapt to at least 4°C of warming. Given the severity of the mitigation challenge that we have described, this seems at present to be a wise precaution.

What would be the cost of such adaptation? The UN Framework Convention on Climate

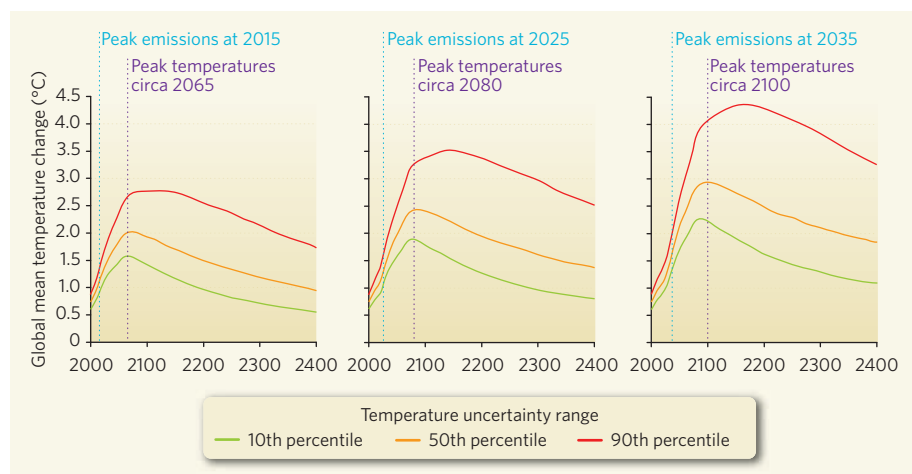


Figure 1 | Temperature scenarios. Global average surface temperature scenarios for peak emissions at three different dates (2015, 2025 and 2035) with 3%-per-year reductions in greenhouse-gas emissions.

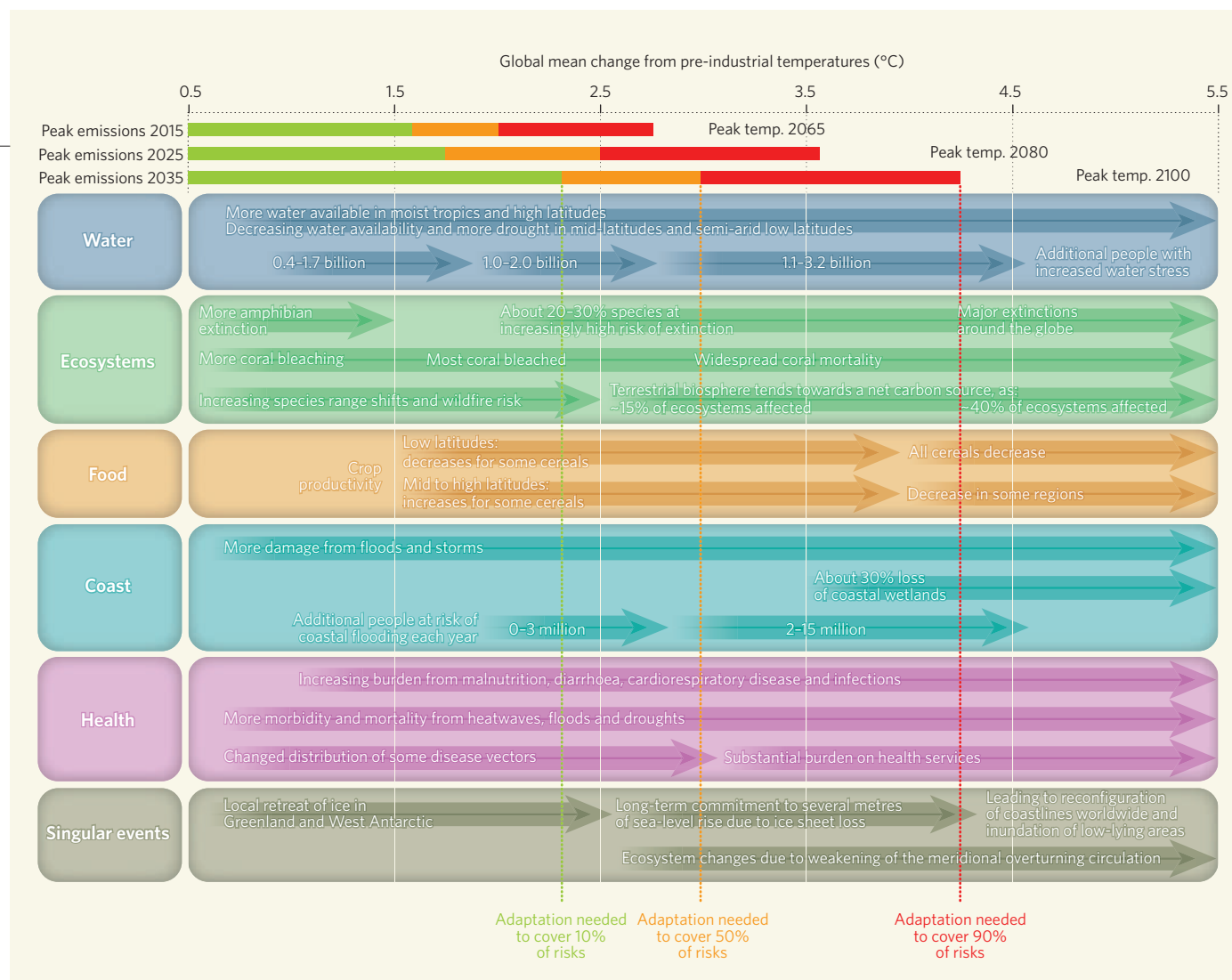


Figure 2 | Expected effects on a range of global sectors for different global mean temperature increases from preindustrial levels. Plotting peak temperatures from the three scenarios explained in Fig. 1 shows the range of projected damaging effects and the adaptation needs for 10%, 50% and 90% coverage of impact risk, and examples of these effects for a range of global sectors (source, ref. 4).

Change (UNFCCC) has estimated that between US\$50 billion and US\$170 billion per year (in current values) will be needed by the year 2030 (ref. 4). This is only a twentieth of current spending on development of new infrastructure globally and a tenth the expected cost of emissions reduction⁵. Such a low figure for the cost of adaptation might lead one to doubt whether climate change poses much of a challenge. The ultimate cost will probably be several times the UNFCCC estimate, however, and much more than this if emissions reduction is delayed or if we wish to protect against high-end uncertainty³.

Additionally, much adaptation may not be physically possible or economically worthwhile. One estimate is that this impracticable adaptation would amount to two-thirds of all damages — about \$1 trillion per year in 2030 (ref. 6), or ten times the UNFCCC estimate for total adaptation funding. This includes damages to irreplaceable biological systems such as coral reefs or the costs of continuing to irrigate for farming in drying regions.

A final concern is that the multi-century-

long recovery process, from peak temperatures this century to roughly 1°C of warming far in the future, may be far from linear. Sea levels, for example, may continue to rise for some decades after land areas have begun to cool; and there is also the possibility that extended melting in the Arctic would reduce albedo and increase methane release, pushing the warming peak higher and further into the future⁷.

The window of opportunity for beginning effective long-term action on climate change is extraordinarily narrow. Urgent and major emissions reductions are essential to avoid the most severe effects. Yet even the most prompt and stringent action still risks overshooting a target of 2°C, and it will require centuries to achieve a roughly stable climate with tolerably low amounts of warming. The consequent demands on adaptation will be enormous, many times those currently envisaged. We should therefore give policies of adaptation much more urgent attention.

Martin Parry was co-chair of the IPCC's 2007 Working Group II assessment and is now at the

Grantham Institute for Climate Change and Centre for Environmental Policy, Imperial College London, London SW7 2AR, UK.

Jason Lowe is head of mitigation advice at the Met Office, Exeter EX1 3PB, UK.

Clair Hanson was deputy director of the 2007 IPCC Working Group II technical support unit and is at the University of East Anglia, Norwich NR4 7TJ, UK.

e-mail: martin@mlparry.com

See also Editorial, page 1077, and www.nature.com/climatecrunch

1. European Commission *Limiting Global Climate Change to 2 degrees Celsius* (2007).
2. Lowe, J. A. et al. *Environ. Res. Lett.* **4**, 014012 (2009).
3. IPCC *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds. Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) (2007).
4. UNFCCC *Investment and Financial Flows to Address Climate Change* (2007).
5. Stern, N. *The Economics of Climate Change: The Stern Review* (Cambridge Univ. Press, 2007).
6. Ackerman, F., Stanton, E. A., Hope, C. & Alberth, S. *Energy Policy* (in the press).
7. Parry, M. L., Lowe, J. A., Palutikof, J. & Hanson, C. E. *Nature Reports Climate Change* doi:10.1038/climate.2008.50 (2008).

ESSAY

The worst-case scenario

Stephen Schneider explores what a world with 1,000 parts per million of CO₂ in its atmosphere might look like.

Thinking about worst-case scenarios is nothing new — climate scientists have been doing it for more than 20 years. In 1988, after intense heat waves baked the eastern and central United States, Robert Watson, later to chair the Intergovernmental Panel on Climate Change (IPCC), and I briefed Bill Bradley, the Democrat senator for New Jersey, on the risks of disproportionate surprises from rapid, major climate change. The nature of those surprises was then, as it is now, unclear in details, although we had our hunches.



What is new is the assertion that we know the level of warming required to pass tipping points for potentially irreversible outcomes — for example, the risk of unstoppable ice sheet melt in Greenland¹. In truth, we don't know the precise values for tipping points, but we can reasonably estimate with medium confidence by looking at palaeoclimates and recent ice sheet behaviour². For Greenland, I estimate, after listening to expert judgements, a few per cent chance that meltwater transporting heat downward has already begun to obliterate ice cover irrevocably. At 1°C more, I'd up my odds to maybe 25% and at 2°C to 60%. At 3°C, because the system is highly non-linear, to 90%. Deficiencies in current knowledge allow us to make only subjective probabilistic estimates that must be revised with new knowledge.

But what if the worst-case scenario came to pass? An atmosphere in 2100 with 1,000 parts per million of carbon-dioxide equivalent would be catastrophic. To understand the effect of this, we need to peer into what Harvard University economist Marty Weitzman calls the 'fat tail'³ of the probability distribution for climate damage. Although the likelihood is uncertain — and probably low — we should give these events more attention because not doing so could be potentially disastrous.

An unthinkable scenario?

In 2000, the IPCC published its *Special Report on Emissions Scenarios*, a now-famous set of storylines for future greenhouse-gas emissions. Even its most optimistic scenario projected a doubling of pre-industrial CO₂ levels by 2100 — this is a scenario that the authors called B1, with emissions-reducing technologies spreading throughout a world with low population

growth and a more egalitarian distribution of resources. At the other end of the spectrum was a 'fossil intensive' scenario called A1FI⁴. This tripled CO₂ to roughly 950 p.p.m. by 2100. I describe this scenario as business as usual, with economic growth deemed more important than conservation.

Often studies of climate change use B1 and A1FI as 'bookends' to bracket future projections. The authors of the *Special Report on Emissions Scenarios* were unable to agree which scenario was most probable and deemed them all "equally sound". Recent history, however, suggests otherwise. Until the economic downturn in late 2008, actual emissions since 2000 have been above the worst-case A1FI scenario⁵. Of course, short-term trends — either above or below long-term scenarios — cannot reliably be extrapolated. Nevertheless, if we resume the pattern of the recent decade, emissions will be on track to making 1,000 p.p.m. of CO₂ more probable than the B1 storyline.

How would 1,000 p.p.m. translate into temperature changes? The amount of global warming associated with any level of radiative energy added to the Earth-atmosphere system — called 'radiative forcing' — depends on the 'sensitivity' of the system. Sensitivity is a measure of how much the surface will warm up if CO₂ levels double from pre-industrial levels, and the IPCC has estimated it to have a "likely" range (implying a 66–90% probability) of 2°C to 4.5°C, with a "best guess" median of 3°C. That implies that there is a 5–17% chance of warming above or below those endpoints. The IPCC estimates about 2.5°C to 6.4°C as the "likely" range for warming by 2100 under A1FI, so there is a 5–17% chance that temperatures will go up by more than 6.4°C by 2100.

Many will argue that warming above 6.4°C is unthinkable. Unfortunately, when I talk to analysts or economists such as Weitzman, I am told that it is precisely the warmer endpoints that they want us to examine further to alert society to catastrophic outcomes that are more than 5–10% likely to happen. This is a probability that is way above the threshold at which people usually buy insurance, or for department of defence deterrence strategies.

The IPCC's Fourth Assessment Report attempted to assess the greatest risks posed by climate change. It suggested that the five

"reasons for concern" examined in the Third Assessment Report remained a valid way to approach risks. But the temperature thresholds at which such damages might be triggered had to be lowered. The figure, published independently by IPCC authors after the report was approved, illustrates this evolution of author judgement⁶. It also extends the possibility of warming to 7°C.

What could be lost

In a 1,000 p.p.m. scenario, many unique or rare systems would probably be lost, including Arctic sea ice, mountain-top glaciers, most threatened and endangered species, coral-reef communities, and many high-latitude and high-altitude indigenous human cultures.

People would be vulnerable in other ways too: Asian mega-delta cities would face rising sea levels and rapidly intensifying tropical cyclones, creating hundreds of millions of refugees; valuable infrastructure such as the London or New York underground systems could be damaged

or lost; the elderly would be at risk from unprecedented heat waves; and children, who are especially vulnerable to malnutrition in poor areas, would face food shortages.

Fairness must also be taken into account, given that some

people would be at much greater risk than others: poor people in hot countries with little adaptive capacity, for instance, indigenous peoples and those exposed to hurricanes or wildfires, or living in low-lying areas. The elderly and children with asthma or other lung ailments would be particularly affected by urban air pollution or wildfire smoke plumes exacerbated by the extreme warming.

The economic outlook is no better. With warming of just 1–3°C, projections show a mixture of benefit and loss. More than a few degrees of warming, however, and aggregate monetary impacts become negative virtually everywhere; and in a 1,000 p.p.m. scenario current literature suggests the outcomes would be almost universally negative and could amount to a substantial loss of gross domestic product. Millions of people at risk from flooding and water supply problems would provide further economic challenges⁷.

The number and intensity of abrupt events and the possibility of irreversible damages goes up non-linearly with warming. If CO₂

"We have to do a lot of things as part of a climate-energy policy portfolio."

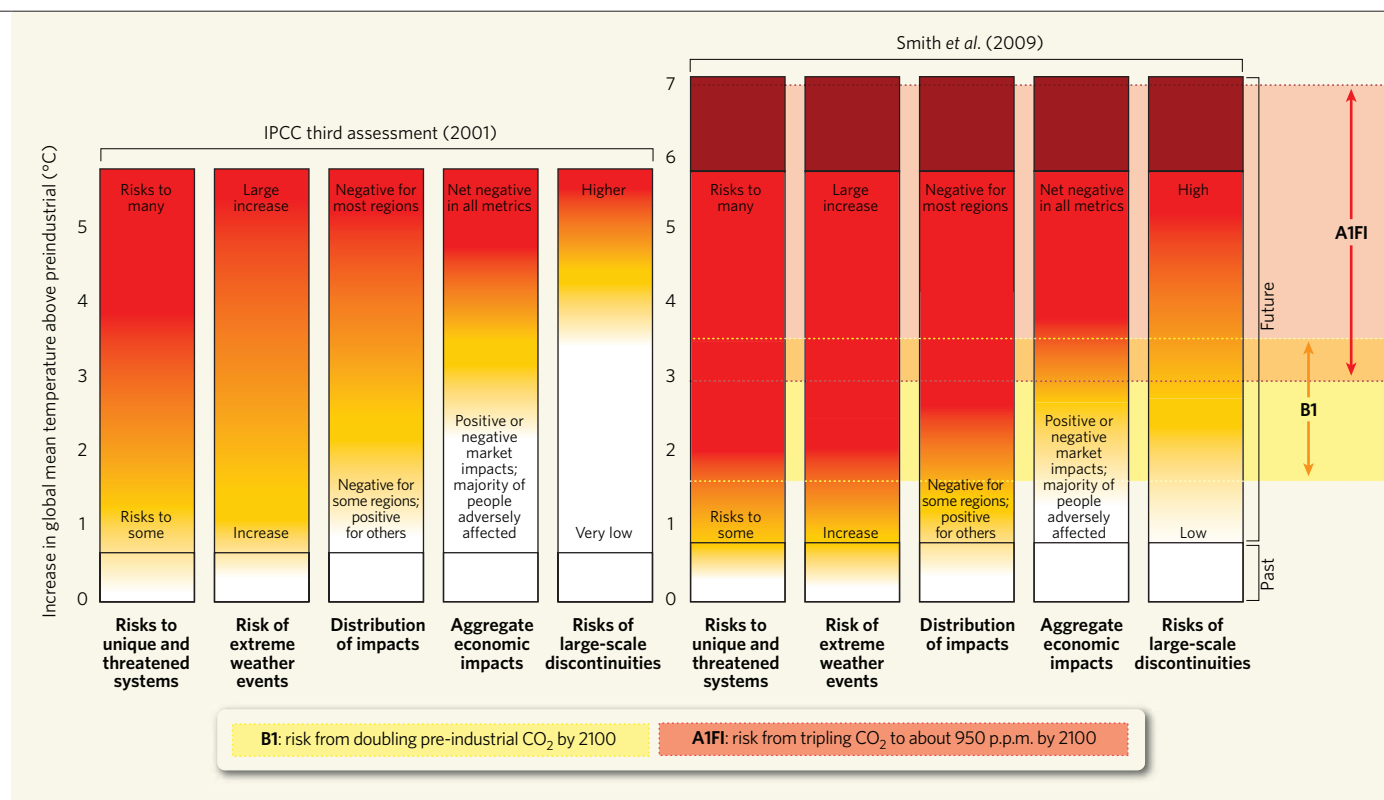


Figure 1 | Updating the embers. The Intergovernmental Panel on Climate Change assessed five reasons for concern in terms of societal, economic and natural damage that would be caused by climate change. The result was the ‘burning embers’ diagram, first seen in 2001. Updates to judgements about the thresholds at which such damages might occur revised the thresholds downwards⁶. Plotting the range of temperature increases caused by a doubling of carbon-dioxide levels by 2100 (B1) and a tripling (A1FI) offers a range of risk levels associated with these scenarios for the five reasons for concern. A 1,000-parts-per-million scenario would look slightly higher than A1FI.

levels were to reach 1,000 p.p.m., a rise in sea levels of up to 10 metres after many centuries from the melting of the Greenland and West Antarctic ice sheets would be more likely². So would damage to coral and oceanic phytoplankton, as their calcium carbonate skeletons could dissolve in acidified oceans. Tropical rainforests would become more vulnerable to wildfire, and in some models such forests would switch from CO₂ sinks to sources, adding yet more emissions. Extinction of some half of known plant and animal species would become much more likely, particularly if climate sensitivity is in the middle-to-upper part of the bell curve^{2,8}.

Responses to 1,000 p.p.m.

An atmosphere with 1,000 p.p.m. CO₂ would produce a rapidly multiplying set of interconnected risks and would undoubtedly spur calls for geoengineering schemes to try to offset the worst effects. But how effective would these schemes be? Injecting dust into the stratosphere to reflect light and prevent radiative forcing, for example, would not prevent increasing ocean acidification.

Similarly, because anthropogenic CO₂ concentration increases and the associated

warming are predicted to last for a millennium or so⁹, geoengineering — and the global cooperation it would require — would have to be sustained without interruption from wars or other political stresses. Moreover, severe unexpected climatic events during a period when climate control was in practice could lead to unprecedented liability claims¹⁰.

It seems obvious that we need to stay well below 1,000 p.p.m. A whole range of policies will be needed, with international cooperation, and they cannot wait for the outcome of the current political negotiations. These include policies that discourage polluting technologies and provide incentives for using cleaner ones, as well as penalties for non-compliance. “There’s no silver bullet, but there’s a lot of bronze buckshot” is something of a cliché among those determined to address climate policy. In other words, we have to do a lot of things as part of a climate-energy policy portfolio. Even if governments debate the emissions-cap target or the distribution of burdens for paying a price for carbon for several years more, at least by rapid implementation of performance standards and clean-technology development we can get on with the job in a politically less contentious way. ■

Stephen H. Schneider is professor of interdisciplinary environmental studies and biology, and a senior fellow in the Woods Institute for the Environment at Stanford University, Stanford, California 94305, USA.
 e-mail: shs@stanford.edu

See also Editorial, page 1077, and www.nature.com/climatecrunch

- Hansen, J., et al. *Atmos. Chem. Phys.* **7**, 2287–2312 (2007).
- Schneider, S. H. et al. in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) Ch. 19, 779–810 (Cambridge Univ. Press, 2007).
- Weitzman, M. L. *The Review of Economics and Statistics* **91**, 1–19 (2009).
- Nakicenovic, N. et al. *Special Report on Emissions Scenarios* (Cambridge Univ. Press, 2000).
- Raupach, M. R. et al. *Proc. Natl Acad. Sci. USA* **104**, 10288–10293 (2007).
- Smith, J. B. et al. *Proc. Natl Acad. Sci. USA* **106**, 4133–4137 (2009).
- Kundzewicz, Z. W. et al. in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) Ch. 3, 173–210 (Cambridge Univ. Press, 2007).
- Fischlin, A. et al. in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) Ch. 4, 211–272 (Cambridge Univ. Press, 2007).
- Solomon, S., Plattner, G.-K., Knutti, R. & Friedlingstein, P. *Proc. Natl Acad. Sci. USA* **106**, 1704–1709 (2009).
- Schneider, S. H. *Phil. Trans. R. Soc. Lond. A* **366**, 3843–3862 (2008).

SPRING BOOKS

Could climate change capitalism?

Economist Nicholas Stern's latest book is a rare and masterly synthesis of climate-change science and economics. His 'global deal' could change capitalism for the better, says **Robert Costanza**.

A Blueprint for a Safer Planet: How to Manage Climate Change and Create a New Era of Progress and Prosperity

by Nicholas Stern

Bodley Head: 2009. 256 pp. £16.99



When economist Nicholas Stern released his 700-page review of the economics of climate change for the UK government in 2006, it fundamentally

reoriented discussions of the subject. The review opened the way for a broader, more realistic and more relevant approach than economists had provided until then.

In his new book, *A Blueprint for a Safer Planet*, Stern takes things further and lays out a roadmap for managing the climate crisis. He devotes much of the book to describing basic characteristics of the climate problem, including our current scientific understanding of the Earth system and humanity's role within it, the dangers posed to human societies by climate disruption, and the inherent uncertainties of climate change and how we can best deal with them. He raises thorny but unavoidable ethical issues, such as how we should weigh up the possible costs and benefits of climate change now and in the future — known as discounting. He explains what policies we should adopt to reduce greenhouse-gas emissions and what we can learn from current good practice in reducing them.

Although much of this material is well known, Stern presents an up-to-date, logically argued synthesis, using a style that makes his book more intelligible than many others on the topic. For example, he masterfully explains the advantages and disadvantages of two of the main mechanisms proposed for limiting emissions. One is a carbon tax that would charge

polluters based on the emissions they produce. The other is 'cap, auction and trade', in which a global emissions limit, or cap, is internationally agreed, permits to emit are auctioned, and holders then buy and sell the permits.

Stern clearly describes the trade-offs between the price certainty of a tax and the quantity certainty of a cap system, emphasizing that "we cannot have both price and quantity certainty in an uncertain world". He also points out some less obvious characteristics of each option, such as how caps "allow international private-sector flows of carbon finance from rich to poor countries", and how they might deal better with the oligarchic, price-manipulating nature of the oil and gas industry than a tax would. He concludes that a mix of policies will be necessary,

but that a clear understanding of the inherent trade-offs is needed to optimize this mix.

Stern's treatment of the issue of discounting the future is also exemplary. The main criticism of the *Stern Review on the Economics of Climate Change* came from a few economists who disagreed with the low discount rates it used to handle this issue. Discounting is about how much weight we should give to costs and benefits that occur in the future, relative to the present. The high discount rates of around 3–6% used by some economists may be useful for comparing small-scale public investments, such as bridges or roads, but are totally inappropriate for issues such as climate change. At 6%, any impacts that might occur more than 50 years in the future are negligible. Stern clearly



ILLUSTRATIONS BY ANNABEL WRIGHT

NEW IN PAPERBACK



Supercontinent: Ten Billion Years in the Life of Our Planet

by Ted Nield (Harvard Univ. Press, \$18.95)

Geologist Ted Nield gives a thorough account of Earth, from long before Pangaea to far into the future. "To handle it without oversimplification or getting lost in a maze of detail is no small accomplishment," wrote David Oldroyd in his review of the hardback edition (*Nature* **449**, 540; 2007).



Earth Under Fire: How Global Warming is Changing the World

by Gary Braasch (Univ. California Press, £14.95)

Award-winning photojournalist Gary Braasch supplies breathtaking imagery of the effects of climate change. He includes personal accounts from eyewitnesses and researchers, together with the best evidence available, to give a refreshing and intelligent take on this well-covered field.

describes the ethical basis for discounting, and puts to rest criticisms of the low discount rates used in the *Stern Review*.

The essence of *Blueprint for a Safer Planet*, however, is the chapter in which Stern outlines the structure of a proposed global climate deal. He bases his argument on three guiding principles: effectiveness, efficiency and equity. His suggestions are complex and interrelated, and he emphasizes that they are an integrated package, not a menu from which selections can be made. Stern also stresses that no deal will work without true international collaboration.

His proposal focuses on six elements: targets for emissions from rich countries and world emissions; targets for developing countries; an effective international emissions-trading regime; combating deforestation; technological advances to reduce emissions; and overseas assistance to help developing countries adapt to climate change.

Stern proposes global targets for cutting emissions by half by 2050, relative to 1990 levels, with developed countries acting first and more aggressively with cuts of around 20–40% by 2020 and 80% by 2050. Developed countries would demonstrate the feasibility of 'low-carbon growth', share technologies with developing countries and implement trading and financial mechanisms. Based on the successful example of these countries, developing countries would commit to targets by 2020, reaching a target of 2 tonnes of carbon per capita by 2050.

Stern's global deal contains all the right elements. We need targets, but as we learn more we may discover that even bigger cuts are required to avoid major climate impacts. We will need enforceable caps coupled with a rate of decline that can be adjusted over time. We must also combat deforestation, effectively spread new technologies and directly assist developing countries with adaptation, but the challenge is how to do all this in a coordinated, focused and sustainable way.

Stern recognizes that this will be costly and complicated to achieve in practice. As the former chief economist of both the European

Bank for Reconstruction and Development and the World Bank, he has long experience of multilateral institutions and recognizes that existing ones are not up to the task. Under the current arrangements, there is no single authority that is responsible for ensuring compliance with climate commitments. Similarly, there is no international funding mechanism dedicated to climate change. Although Stern is hesitant to recommend new institutions, which are costly to set up and run, he has learned from experience that "coherence within climate-change activities is of great importance and without a clear primary responsibility in one place it would be very elusive", noting that "priority would slip relative to other shorter-term issues in ... other organizations". The design of this new global institution is not addressed, however.

One institutional design that might meet all of

"A global deal for climate change has positive implications for global capitalism."

Stern's criteria, as well as alleviate poverty, is the 'Atmospheric Trust' proposed by Peter Barnes and others (P. Barnes *et al. Science* 319, 724; 2008). Such a trust would use a cap, auction and trade system, reducing the cap over time to stabilize greenhouse-gas

concentrations at the desired target. A fraction of the revenues would be returned as 'dividends' on a per-capita basis to everyone on Earth. This would directly address poverty issues. The rest would be invested in protecting and enhancing the atmospheric asset — as proposed in Stern's global deal — using initiatives such as payments for the carbon-sequestration services of forests to combat deforestation, investment in open-access renewable-energy technology to reduce emissions and direct assistance to developing countries for adaptation to climate change.

One fundamental shortcoming of the book, however, is its uncritical acceptance of economic growth as the only path to future prosperity. Stern acknowledges that any future growth must be 'low carbon', but fails to recognize that conventional economic growth is merely a means to the goal of sustainable human well-being. Economic growth is not — and should not be — an end in itself.

There is evidence in developed countries that economic growth beyond a certain point does not improve well-being, owing to the hidden, external costs of that growth, including climate impacts. For example, an oil spill increases gross domestic product (GDP) as someone must pay to clean it up, yet it detracts from well-being. Increased crime, sickness, war, pollution, fires, storms and pestilence are all positive for GDP because they increase economic activity. We need to move beyond GDP as a measure of well-being — something for which it was never designed — and develop and use better indicators of sustainable quality of life. Proposed alternatives include the Genuine Progress Indicator and Gross National Happiness, as used in Bhutan in south Asia, but we need to build a broad global consensus on alternative measures to move forward.

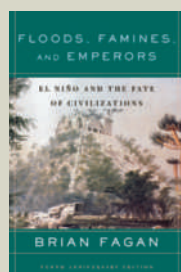
Stern may not fully recognize that a global deal for climate change has positive implications for global capitalism. The deal will require a new, more nuanced suite of property rights and responsibilities that give adequate weight to 'the commons' — public goods such as the atmosphere that are open-access but need to be assigned appropriate property rights so they can be protected. We cannot, and should not, assign private property rights to this inherently common asset, but a global institution that charged for greenhouse-gas emissions would at least implicitly assign such common rights.

The key is to find the right balance between private and common property while protecting the commons. Neither socialism, in which most property is common, nor capitalism, in which most is private, have dealt adequately with open-access commons because they have failed to get this balance right. The meltdown of these economic systems presents the opportunity to find a new balance that will help us lay the path to sustainable prosperity. ■

Robert Costanza is director of the Gund Institute for Ecological Economics at the University of Vermont, Burlington, Vermont 05405, USA. e-mail: robert.costanza@uvm.edu

See Editorial, page 1077.

Listen to an interview with Nicholas Stern at www.nature.com/nature/podcast.



Floods, Famines and Emperors: El Niño and the Fate of Civilizations (Tenth Anniversary Edition) by Brian Fagan (Basic Books, \$17.95)

First published in 1999, Brian Fagan's book charts the discovery of El Niño — the Pacific ocean-atmosphere oscillation underlying freak weather events — and shows how climate change affected ancient civilizations. In describing how they coped, or not, Fagan highlights the problems we face in dealing with climate change today.



American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer by Kai Bird and Martin J. Sherwin (Atlantic Books, £9.99)

This impressively researched and well-written book explores the life of nuclear physicist J. Robert Oppenheimer, covering the different sides of his personality and his rise and fall in society. Bird and Sherwin provide a thorough exploration of the science and politics of the nuclear age.

Why inequality is fatal

The Spirit Level: Why More Equal Societies Almost Always Do Better

by Richard Wilkinson and Kate Pickett

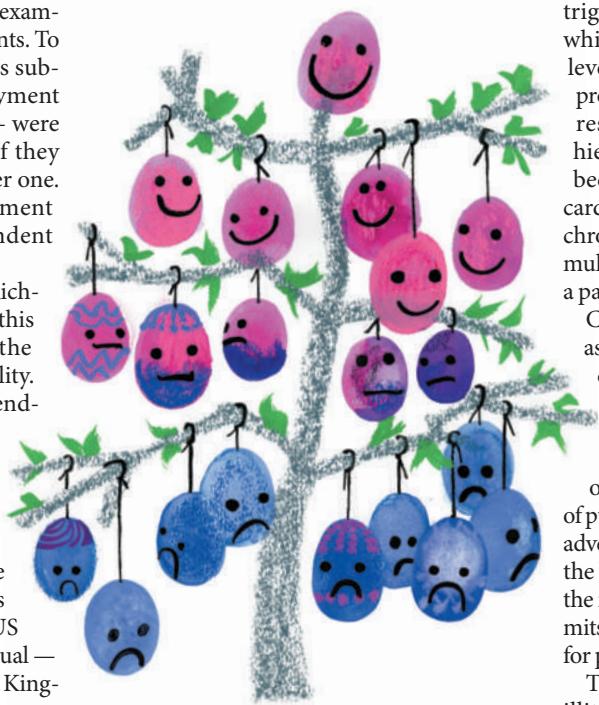
Allen Lane: 2009. 320 pp. £20

Why are our chances of reaching a great age so affected by wealth and status? The obvious answer is that more income buys better health. But it is a lot more subtle than that, as shown three decades ago by the Whitehall Study, in which epidemiologist Michael Marmot examined the death rates of British civil servants. To the surprise of many, he found that his subjects — all in continuous paid employment and with equal access to health care — were more likely to die in any given year if they were in a lower-grade job than a higher one. Marmot concluded that the employment hierarchy itself created status-dependent stress that affected the workers' health.

In their new book, epidemiologists Richard Wilkinson and Kate Pickett extend this idea with a far-reaching analysis of the social consequences of income inequality. Using statistics from reputable independent sources, they compare indices of health and social development in 23 of the world's richest nations and in the individual US states. Their striking conclusion is that the societies that do best for their citizens are those with the narrowest income differentials — such as Japan and the Nordic countries and the US state of New Hampshire. The most unequal — the United States as a whole, the United Kingdom and Portugal — do worst.

Many measures of the quality of life, including life expectancy, are correlated with the degree of economic equality in each country. A variety of problems such as mental illness, obesity, cardiovascular disease, unwillingness to engage with education, misuse of illegal and prescription drugs, teenage pregnancy, lack of social mobility and neglect of child welfare increase with greater inequality. Violence, from

murder to the bullying of children at school, follows the same pattern. These trends are tied up with issues of trust: the authors chart a profound decline in trust in the United States from the 1960s to the present, which matches rising inequality during the long Republican ascendancy. The only statistic that defies the general trend is that for suicide, the incidence of which is higher in the most egalitarian countries, such as Sweden and Japan. Those



who feel out of step with their country seem to blame themselves and not others.

The idea that income inequality within a society is more unsettling to health and welfare than income differences between societies has been hotly debated for more than two decades. In the past year alone, six academic analyses have been published in peer-reviewed journals,

four of which contradict the hypothesis on statistical grounds. Yet Wilkinson and Pickett do not address these criticisms in their book. They might also have explained the occasional notable deviation from their theory, such as the unexpectedly high murder rates in egalitarian Finland and the unexpectedly low rates in very unequal Singapore.

How can inequality affect such a diverse set of social problems so profoundly? The authors make a compelling case that the key is neuroendocrinological stress, provoked by a perception that others enjoy a higher status than oneself, undermining self-esteem. This triggers the release of the hormone cortisol, which raises blood pressure and blood sugar levels, from which myriad health and social problems unfold. This seemingly hard-wired response has been well studied in social hierarchies of monkeys; low-status animals become predisposed to atherosclerosis and cardiovascular disease. Humans experiencing chronic stress exhibit similar symptoms, accumulating abdominal fat under the influence of a part of the brain associated with addiction.

Cortisol overrides 'feel-good hormones' such as oxytocin, involved in establishing trust, and dopamine, the reward signal that reinforces memory, attention and problem-solving ability. Cortisol-induced stress predisposes some individuals to mental illness or violent behaviour. It can hasten the arrival of puberty, which may prompt premature sexual adventures, providing a plausible explanation of the high prevalence of teenage pregnancies in the most unequal societies. Cortisol also transmits stress to a fetus, with lasting consequences for physical and emotional development.

The stress response could even exacerbate illiteracy and unwillingness to engage with education. Wilkinson and Pickett argue that these are more common in less equal societies, not because of poverty but because school-age students may lose self-esteem when they realize that some of their peers are better equipped than themselves for educational challenges. The stress response may also lead to illicit drug use. Monkey social hierarchies provide a clue: dominant animals secrete dopamine and feel

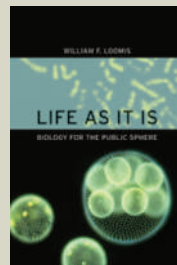


The Secret History of the War on Cancer

by Devra Davis

(Basic Books, £10.99)

Drawing on research and personal experience, Devra Davis argues that focusing on cancer treatment instead of prevention has cost lives. Reviewer Daniel S. Greenberg wrote, "For a well-documented, prosecutorial account of the dark side of cancer-control politics, Davis's work — lopsided and verbose as it is — merits attention." (*Nature* **449**, 660–661; 2007).



Life As It Is: Biology for the Public Sphere

by William F. Loomis

(Univ. California Press, \$15.95)

William F. Loomis explores controversial issues from a biological perspective, providing "a fascinating, if occasionally disjointed, survey of topics ... the nature and evolution of life, and current scientific thought regarding consciousness, psychology and social behaviour." (Eugenie Scott, *Nature* **452**, 690–691; 2008).

good about their place in the world, whereas monkeys at the other end of the status scale are more inclined to self-medicate — with cocaine if given the opportunity.

The Spirit Level is a brave and imaginative effort to understand the intractable social problems that face rich democratic countries. For Wilkinson and Pickett, economic equality is the best way to improve the quality of life for all. Governments can get there by using redistributive taxation and an extensive welfare

state, as in Sweden, or by restraining income disparities and minimizing public spending, as in Japan. The book ends optimistically: whatever route is chosen, the authors argue, the current economic slump may be a providential opportunity to start righting the balance. ■

Michael Sargent is a developmental biologist at the National Institute for Medical Research in London and is author of *Biomedicine and the Human Condition: Challenges, Risks and Rewards*. e-mail: msargent@nimr.mrc.ac.uk

Fiction beyond the grave

Sum: Forty Tales from the Afterlives

by David Eagleman

Pantheon: 2009. 107 pp. \$20.00

There is no life after death. When our bodies fail, our minds go with them, and the game is over. Or is it?

Sum gathers 40 playful sketches of what an afterlife might hold for us, from expanding into a nine-dimensional cloud to working as an extra in other people's dreams. As rigorous and imaginative as the writings of Italo Calvino and Alan Lightman, each vignette is a glimpse into an expansive topic such as time, faith or memory. Together they illuminate an astounding range of possibilities for the meaning of human life.

Neuroscientist David Eagleman has written these fictional scenarios in parallel with a number of his popular science books about the brain, and while running a laboratory at Baylor College of Medicine in Houston, Texas. His research on time perception may have inspired some of his literary conceits. In the title story, a lifetime of activities is sorted into insufferable batches — 30 years of sleep, 200 days in the shower, 18 days staring into the refrigerator — suggesting that it is the transitions between experiences that make life worth living.

The book includes, as one might expect, a round of fables that deflate Christian stereotypes of the hereafter. In some, paradise is

vulnerable to the petty vices of men — holy war, bickering, bureaucracy, even communism — which makes these versions of heaven more like comic varieties of hell. In another, God is revealed to be an opportunistic tinkerer who doesn't understand His own creation.

Eagleman is at his sharpest when he envisions efforts to evade death using science. In one tale, a doctor rids the world of mortality only to be killed by rioters nostalgic for natural death. In another, the elderly pay a company to upload their minds into computers that would stimulate them with their own private afterlives



for eternity — if only the machines worked. Death is an essential part of life.

In other stories we are the malfunctioning machines, built by another species to map Earth, perform a computation or discover the meaning of life. The engineering project usually fails, but humanity flourishes nonetheless. The moral seems to be that we should look for the silver lining in our own design flaws.

Sum becomes unsettling when it turns to cosmology. A civilization, after discovering their universe to be alive, attempts to communicate with her. "We sent a sharply defined sequence of electromagnetic pulses," its citizens say, "which interacted with local magnetospheres, which influenced asteroid orbits, which nudged planets closer and farther from stars, which dictated the fate of lifeforms, which changed the gases in the atmospheres, which bent the path of light signals, all in complex interacting cascades" that "took a few hundred years... to arrive at her consciousness". The cosmic answer, perhaps nothing more than an immune response, is the destruction of their planet, leading a survivor to conclude that "communicating with [the universe] is not impossible, but it is pointless".

Hope returns when Eagleman trades in his telescope for a microscope. There is some comfort in the idea that "when you die, you are grieved by all the atoms of which you were composed".

The best stories in *Sum* remind us that it is natural to want to know our place in the scheme of things. The book is a scripture of sorts, but because each myth contradicts the last, it is not a dogmatic collection. Eagleman has said that he is neither a believer nor a non-believer in the conventional religious sense. Rather, he considers himself a "possibilian", which he defines as a creed for "those that celebrate the vastness of our ignorance, are unwilling to commit to any particular made-up story, and take pleasure in entertaining multiple hypotheses". These may indeed be the qualities of a good scientist — and a good storyteller. ■

Jascha Hoffman is a writer based in New York. e-mail: jascha@jaschahoffman.com

DAVID
McFARLAND

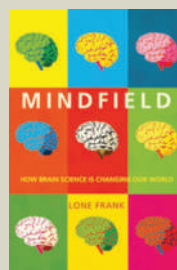
GUILTY
ROBOTS,
HAPPY
DOGS

THE QUESTION OF ALIEN MINDS

Guilty Robots, Happy Dogs: The Question of Alien Minds

by David McFarland
(Oxford Univ. Press, £7.99)

Taking a fresh angle on the question of 'alien minds' — whether animal or machine — and drawing on current research in computing, robotics and animal behaviour, David McFarland offers an accessible introduction to the philosophy of the mind. He explains why intelligence may be impossible to define.



Mindfield: How Brain Science is Changing Our World

by Lone Frank
(Oneworld, \$16.95)

Lone Frank asks how neuroscience is transforming our society. By describing her own experiences while researching the book — from holding half a real brain in her hands to talking with leading scientists — she explains advances in the field for a general audience.

Genes, games and the sexes

Natural selection selects the fittest, but the fittest need not be selfish, according to two new books. Starting from different backgrounds, evolutionary biologist Joan Roughgarden and anthropologist and primatologist Sarah Blaffer Hrdy converge on that message regarding sex and reproduction.

The authors also converge on another point. Theorists go wrong less because of the assumptions they know they are making, and more because of the ones they don't. Roughgarden shows how modellers can agree about the maths and the results of particular models, yet, thanks to rival metaphysical assumptions, still disagree fiercely about the verbal narratives they attach to the models. For Blaffer Hrdy, the problem is simpler: it stems largely from the sex of the researchers. She is interested in primate, ape and human mothers. Early researchers were mostly men, but recently, large numbers of women scientists have been asking new questions and noticing different things.

The Genial Gene is the latest round in the debate between Roughgarden's social-selection theory and sexual-selection theory. Many have contributed to it, but I shall concentrate on Roughgarden's responses to her most thoughtful critic, zoologist Tim Clutton-Brock.

Roughgarden starts by differentiating sex from the terms male and female. Sex is the combining of gametes from two parents. The production of large gametes by females and small ones by males — anisogamy — is thought to be responsible for competition



between the sexes. Since Angus Bateman's fly experiments, females have been portrayed as investing more in their gametes, and therefore

being choosy about their mates. Males invest little in their gametes, mate with as many females as possible and compete with other males to do so. But Roughgarden proposes that anisogamy originally evolved to ensure contact between male and female gametes. Pursuing this to the genetic and cellular levels, she generates an

alternative theory of reproduction.

Roughgarden's approach includes both developmental and evolutionary processes. She also explains reproductive behaviour as a process of bargaining and communication between

the sexes, rather than competition. The two tiers break the genetic determinism implicit in single-tier evolutionary models by granting more plasticity to individuals of both sexes when bargaining. The main innovation, however, concerns the bargaining process itself, and involves game theory.

Game theory traditionally analyses an interaction between two players; in this case, a male and a female. Mathematician John Nash developed its principal theorems in the 1950s, including the concept of Nash equilibria, whereby neither player in a game can do better by changing strategies. John Maynard-Smith introduced game theory to biology — by focusing on fitness pay-offs to interacting individuals in evolving populations, he attempted to capture what happens at both the individual and population levels.

Roughgarden complains that Maynard-Smith introduced only the competitive half of game theory. She introduces the cooperative half, based on organisms communicating, bargaining and allocating side-payments to each other. Cooperation between the sexes leads to Nash bargaining solutions, not to Nash competitive equilibria. Different behavioural solutions at the developmental tier then translate into alternative evolutionary outcomes. As someone who works on the developmental process of niche construction, whereby the actions of organisms generate feedback in evolution by modifying natural selection pressures in environments, I like Roughgarden's two tiers.

That does not mean Roughgarden is correct,

The Genial Gene: Deconstructing Darwinian Selfishness

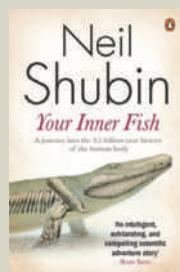
by Joan Roughgarden

University of California Press: 2009. 272 pp. \$24.95, £14.95

Mothers and Others: The Evolutionary Origin of Mutual Understanding

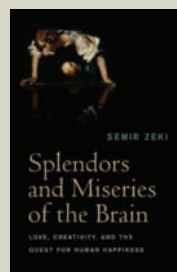
by Sarah Blaffer Hrdy

Belknap/Harvard University Press: 2009. 432 pp. \$29.95, £19.95, €21.00



Your Inner Fish: A Journey Into the 3.5-Billion-Year History of the Human Body

by Neil Shubin (Penguin, £9.99)
Neil Shubin "delves into human gristle, interpreting the scars of billions of years of evolution that we carry inside our bodies", wrote Carl Zimmer in his review (*Nature* **451**, 245; 2008). "The simple, passionate writing may turn more than a few high-school students into aspiring biologists."



Splendors and Miseries of the Brain: Love, Creativity and the Quest for Human Happiness

by Semir Zeki (Wiley-Blackwell, £16.99)
Semir Zeki examines the brain's ability to seek knowledge and form concepts in creative areas such as music, art and literature. He gives us an insight into how the brain functions, what this means for happiness, and how even negative emotions can be a source of creativity.

nor is sexual selection wrong. The awkward data for sexual-selection theory are not yet decisive. Also, as the mutual concessions and careful arguments between Clutton-Brock and Roughgarden show, there is often little empirical ground between them. Sexual selection starts with competition, but admits cooperation as a by-product. Social selection starts with mutual regard and cooperation, but admits competition when bargains break down. To compete, selfish genes must cooperate. To cooperate, genial genes must compete. It is difficult to sort between them. But Roughgarden succeeds in re-opening issues long thought closed.

Blaffer Hrdy's book is narrower in scope but also provocative. She argues that unlike other apes, *Homo sapiens* could never have evolved if human mothers had been required to raise their offspring on their own. Human infants are too helpless and too expensive in their demands for care and resources. So human females have to line up helpers — sometimes extending beyond their own kin — to raise their young. That requires both males and females to invest heavily in social skills for bargaining with other members of their groups. Blaffer Hrdy suggests that females in ancestral hunting and gathering groups may have thrived because they were free to be flexible in this way. Female flexibility was reduced when humans established settlements requiring male coalitions to defend them, probably leading to greater control of females by males.

More rides on these books than the relationship between the sexes. In her last chapter, Roughgarden distinguishes between Charles Darwin's repeatedly verified theory of "descent with modification", versus Herbert Spencer's unverified notion of the "survival of the fittest". It was Spencer who encouraged social Darwinism. Clearly we can go wrong if we attach faulty metaphors and narratives to evolutionary theory. The most refreshing aspect of both these books is the challenges they offer to what we thought we already knew. ■

John Odling-Smee is an emeritus fellow at Mansfield College, University of Oxford, Oxford OX1 3TF, UK.
e-mail: john.odling-smee@anthro.ox.ac.uk

Managing nature as Earth warms

Climate change is transforming the world as we know it. It disrupts biological clocks, pushes species to different latitudes and altitudes and shrinks biological diversity. It also challenges humanity to question its relationship with Earth; global warming is the antithesis of responsible stewardship. We look to science for guidance on how to turn back the hand of humankind, but it can provide only partial answers. Science is imperfect, unfeeling and slow compared with the steady rise of the global thermometer.

Into this mix of responsibilities, human identity and scientific uncertainty comes Anthony Barnosky's new book. Unlike other accounts of climate change, *Heatstroke* looks at the issue from the perspective of plants and animals. He introduces us to the Irish elk, the pack rat and biologists such as Jim Patton, a professor at the University of California, Berkeley — species and people that are helping to reveal the biotic signature of climate change.

Using many case studies, Barnosky explains how worldwide changes in climate are altering the reproductive rates, timing of breeding and living conditions of creatures. He is spot on in his description of new and exciting scientific findings, portraying them in an accessible and compelling way. Such findings include predictions of the existence of climates with no modern counterparts within 100 years, and a careful reconstruction of past climatic effects on small mammal communities. The reader becomes engulfed in a world of scientific discovery, searching through bones, walking transects to record the numbers of animal and plant species, and looking back across history to reveal the influence of climate on life.

This is not a happy book. Barnosky sounds the alarm about the biological effects of climate change, but his gloom and doom message could alienate readers. He repeatedly asserts that climate change will lead to

permanent species loss: "At best, we seem to be witnessing wholesale changes in nearly every ecosystem on Earth. At worst, we may be witnessing the extinction of life as we've known it." The former statement is true, but the latter is extreme. In Yosemite and Yellowstone national parks in the United States, Barnosky reports changes in the composition and locations of species. But these places are not yet experiencing profound species loss. The public needs to understand how climate change is altering life on Earth and that such threats are very serious, but I worry that scientists risk a backlash if their primary message is the worst-case scenario.

Climate is a major determinant of where a species lives and how species interact. Biologists also know that climate change will outpace evolution for a great number of organisms, although perhaps not for bacteria, viruses and some insects. Barnosky describes the result as "like taking a color portrait and rendering it in black and white, or stripping all the harmonic notes out of a symphony". But not all organisms will be affected negatively; some will flourish. We need to figure out if climate change is eroding the species that humans value and replacing them with those that cause harm. And we must identify which species will be most affected, which ones will muddle through and which will rise to prominence. This information will help us to determine what sort of biotic world climate change is creating and what steps we might take to affect that change.

If we heed Barnosky's call to care about the impacts of climate change, we must reduce the greenhouse gases that we emit into the atmosphere and capture those already there. We could also help some species out.

Heatstroke describes the strategy of assisted migration — helping a species to relocate to a place where it might be expected to thrive.

Heatstroke: Nature in an Age of Global Warming

by Anthony D. Barnosky

Island Press: 2009.

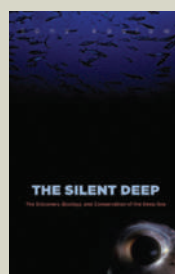
288 pp. \$26.95, £16.99



The Ten Most Beautiful Experiments

by George Johnson
(Vintage, £8.99)

George Johnson's book covers ten of the most historically significant experiments in science — including Isaac Newton's prisms and Ivan Pavlov's dogs. He does justice to each experiment, showing how ingenious and elegant it was — and how the process of experimenting may be as important as the conclusion.



The Silent Deep: The Discovery, Ecology, and Conservation of the Deep Sea

by Tony Koslow

(Univ. Chicago Press, \$22.50)

Describing the huge variety of ocean life "with textbook depth on all aspects of deep-sea science and conservation," (Mark Schroepe, *Nature* 447, 909-910; 2007), Tony Koslow examines how oceanography has developed and discusses human exploitation of the seas.

This strategy aims to overcome dispersal barriers, so that species can survive climate change by altering their geographic distributions. Assisted migration is not a panacea, and it has serious risks, but it is an example of a new kind of thinking that we desperately need. We must devise other adaptation strategies to reduce the harmful effects of climate change where they occur.

Heatstroke concludes with ideas about how to slow the climate crisis, such as reducing energy consumption. But Barnosky stops short of offering suggestions that match the scale of the biodiversity crisis he has outlined. He could have suggested alternative strategies, and I hope that *Heatstroke* will inspire others to design potential solutions. We should also stop confusing assisted migration with 'Pleis-

tocene rewilding', an idea that Barnosky raises after assisted migration. Rewilding would return ecosystems to their state before historical climatic change by transporting large animals and predators across continents. But rewilding has different goals and potentially greater consequences than assisted migration, and it has muddled the waters for productive debates about such strategies.

After reading *Heatstroke*, I felt the urge to go outside and experience nature at first hand, to develop a deeper appreciation for

the life that climate change threatens. If the public at large could be similarly inspired, there might be hope for positive change. Read this book, and reflect on your own views about humanity's place in nature. Then plant a tree, walk to work, and go and call your political representative. ■

Jessica J. Hellmann is professor of biological sciences at the University of Notre Dame, 107 Galvin Life Science Center, Notre Dame, Indiana 46556, USA.
e-mail: hellmann.3@nd.edu



Tales of top models

Pavlov's Dogs and Schrödinger's Cat: Scenes From the Living Laboratory by Rom Harré

Oxford University Press: 2009. 288 pp.
£16.99, \$34.95

This charming book is entertaining, thought-provoking and frustrating. It is worth reading for all three of those reactions.

Written by the distinguished philosopher Rom Harré, *Pavlov's Dogs and Schrödinger's Cat* is based on the premise that just as scientists interrogate the real world with inanimate equipment such as flasks, telescopes and DNA sequencers, we also do so with animate organisms or their parts. Thus, dogs were the apparatus Ivan Pavlov used to study conditioned responses, Galapagos finches the devices used by Peter and Rosemary Grant to study natural selection, and peas and flies the tools used by

Gregor Mendel and Theodosius Dobzhansky, respectively, to elucidate genetics. The use of organic equipment extends beyond the biological sciences: for a time, frogs' legs and canaries were more sensitive detectors of electricity and carbon monoxide than any machine, and we still explore Earth's climatic history using long-departed organisms. Harré romps through five centuries of vignettes, asking who used these organic tools, why and to what end.

And it is an entertaining romp, partly because it is so well written, and partly because the animals and plants involved generate great stories. Some stories are lovely, such as William Buckland's conclusion from cave bones that the UK county of Yorkshire was tropical before Noah's flood. Some are sad, such as the experiments on filial love and maternal separation in monkeys by Harry Harlow, studies that caused great distress to the animals involved, and that have not stood the test of time. And

some are horror stories — including one of the most gripping and succinct summaries of the Lysenko fiasco in Soviet genetics I have read. The scientific questions that lie behind Harré's stories are often startling; for instance, what are the psychological underpinnings of those who commit genocide? The experimental ingenuity of scientists is sometimes breathtaking, such as Stephen Hales' determination of blood pressure using wax, brass pipes, glass tubes and expiring horses. Despite the brevity of each case study, the scientists emerge as real people, with varying degrees of brilliance and interpersonal skills. As a lucid illustration of the messy, chaotic and glorious professional world we scientists have always inhabited, this book is great.

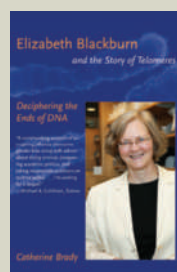
Yet the book is more thought-provoking than a simple storytelling exercise. It is organized not as a history, nor by scientific subject, but by the principles of the philosophy of science. Chapters are devoted to organisms as detecting and measuring devices, and as tools for exploration, for testing hypotheses and for modelling reality. This organization generates



Deadly Companions: How Microbes Shaped Our History

by Dorothy H. Crawford
(Oxford Univ. Press, £8.99)

From the origin of the first microorganisms on Earth, Dorothy Crawford describes how microbes have evolved alongside humanity. Showing how they have altered human history, such as through plague or famine, she explains how humanity has affected microbes in return — and why this will never change.



Elizabeth Blackburn and the Story of Telomeres: Deciphering the Ends of DNA

by Catherine Brady
(MIT Press, \$15.95)

This compelling tale describes the science and politics behind molecular biologist Elizabeth Blackburn's great discoveries. She "emerges as a valuable role model in the sometimes unsettling treatment of women in the world of science", wrote Maria A. Blasco (*Nature* **450**, 613–614; 2007).

an easily digested introduction to many of the key concepts of the philosophy of science. It demonstrates vividly that there is no single way of doing science; philosophers of science are hard pressed to describe what is going on, let alone prescribe what we ought to be doing.

There is another layer too. Harré is obviously troubled by the ethical dilemmas associated with the use of live organisms in science. He argues, quite rightly, that it makes no sense to hold strong views on animal experimentation without having a sound understanding of how science works, of how and why animals have been used, and the unpredictable gains that can result. The book does a terrific job of generating that understanding. Even the inclusion of examples that, at first glance, have little to do with animal experimentation — plants, worms on mounds of methane ice, and virtual animals such as the titular cat and Richard Dawkins' biomorphs — bring into sharp focus issues that are central to the political debate. Harré discusses the merits of inanimate and non-sentient alternatives, and the unpredictable value of generalizing from model organisms.

This book should be compulsory reading for activists who man the barricades, throw the bombs or step up to the microphone. As Harré says, "Only when we have a clear idea of what has been done by whom and for what purpose can we take up the pressing moral questions that must arise."

But then comes frustration. Having provided that "clear idea", Harré doesn't take up the pressing moral questions. He concludes that there are three dimensions to argue about: the extent to which gaining scientific knowledge is an absolute good, the extent to which other living things have inalienable rights, and the extent to which we ascribe mental lives to other things. But then he dodges the bullet, saying "I leave the working out of moral arguments to others more qualified than I am to reach just



and ethically sound conclusions". How much more qualified than Harré can you get?

It is easy to feel that Harré has interesting views that he has not shared. He mentions early on that he has "serious reservations about many projects in which animals have been involved". He hints in two places that experimenting with plants has implications for the animal-experimentation debate. He even says there may be moral issues about working with

virtual organisms. But on none of these does he expand.

Could it be that with animal experimentation, things are so muddy, so difficult, that even a professional philosopher, immersed in the topic and its context, is unable to come to a logically justifiable position? Are the moral issues raised by the scientific use of animals different from those raised by, say, slavery, abortion, euthanasia or the death penalty? They could well be. For the most part, we cannot tell in

advance which areas of scientific knowledge are worth knowing. Nor, for the interesting cases, can we tell whether our instrumentation — animate or inanimate — is up to the game. I think the debate on animal experiments is with us for eternity. ■

Andrew F. Read is professor of biology and entomology at the Pennsylvania State University, University Park, Pennsylvania 16827, USA. e-mail: a.read@psu.edu

A billionaire's vision for India

Imagining India: The Idea of a Renewed Nation

by Nandan Nilekani

Penguin: 2009. 528 pp. \$29.95

Nandan Nilekani's book is a product of the new India. The author, a first-generation, wealthy software entrepreneur, belongs to the iconic trinity of that nation — along with the film star and the cricketer, his words command attention.

On graduating in electrical engineering from the Indian Institute of Technology in Bombay in 1978, Nilekani, unlike many of his classmates, did not emigrate to the United States. In 1981, he and six others pooled US\$250 to start Infosys Technologies, an information-technology (IT) consulting

and service company; its revenue eventually surpassed \$4 billion. Today this company, listed on the US stock exchange and with more than 103,000 employees, has a global footprint.

As Infosys soared, so did the stature of Nilekani. He emerged as one of a new breed of businessmen, blessed by both Lakshmi and Saraswati, the rival Hindu goddesses of wealth and wisdom. He is a role model for young, ambitious Indians. Such is his profile in India that simply emblazoning his face on the cover of the book guaranteed a best-seller.

The various Indian Institutes of Technology, the seedbeds for the technical prowess of Nilekani and his peers, were founded as part of Jawaharlal Nehru's vision. Nehru, India's first prime minister, believed that these state-run institutes, offering quality technical



Flower Hunters

by Mary Gribbin and John Gribbin
(Oxford Univ. Press, \$19.95)

This engaging collection tells the stories of 11 remarkable 'flower hunters' who travelled the globe to discover new plants. It describes the impacts they had on both gardening and science, highlighting the difficulties they experienced on their travels and when trying to propagate the plants they brought back with them.



Starved for Science: How Biotechnology Is Being Kept Out of Africa

by Robert Paarlberg
(Harvard Univ. Press, \$16.95)

In this controversial book, Robert Paarlberg argues that opposition to agricultural science in prosperous countries is reaching Africa, denying poor farmers access to technologies that might improve their yields — especially transgenic crops with insect- or drought-resistance.

education almost for free, would produce the leaders of tomorrow. As Nilekani points out, "India set off on a path of knowledge-intensive growth that was both unique and unusual for a developing economy".

The policies of planned development and state-financed education gave less-privileged children access, first to intellectual capital and later to bank loans. Thus, a strong foundation was laid before some of Nehru's economic policies were reversed in the 1990s.

The author's training in engineering is reflected in the modular structure of *Imagining India*. Nilekani interviews more people than a journalist would and cites as many sources as an academic. The former may see him as a rival in recycling the views of others, and the latter may find him lacking in rigour. But as an IT expert, the author knows a thing or two about 'crowd sourcing' — of information and knowledge.

This book is not the story of Infosys. Nilekani has chosen a broader canvas. He takes in the whole range of the nation's maladies — a frozen mindset, inequity, the halting economic reforms, bumbling bureaucrats, caste-biased sectarian politicians, unethical businessmen, inadequate infrastructure and a crippled higher-education system. His strongest comments are reserved for the failures in the education sector, which lacks competition and quality control and is marked by a proliferation of degree-giving institutions, a shortfall of teachers and a lacklustre administration.

Nilekani laments the sharp polarization of discourse in all of India's spheres — political, economic and social. The mindless 'for or against' clashes that occur trivialize important issues and leave little scope for reasoned debate and policy-making. Such conflicts raise barriers against change and stunt the growth of new ideas. In my view,

such polarization and the weakening social cohesion constitute a far greater danger than a temporary dip in the economic growth rate. However, no fault line shakes the author's optimism.

Nilekani is critical of the early quasi-socialist policies of planned development, state-run enterprises and industrial licensing that stifled initiative and kept entrepreneurs and the economic growth rate shackled. But he is

state-run space, atomic energy and other research programmes also nurtured hundreds of ancillary enterprises.

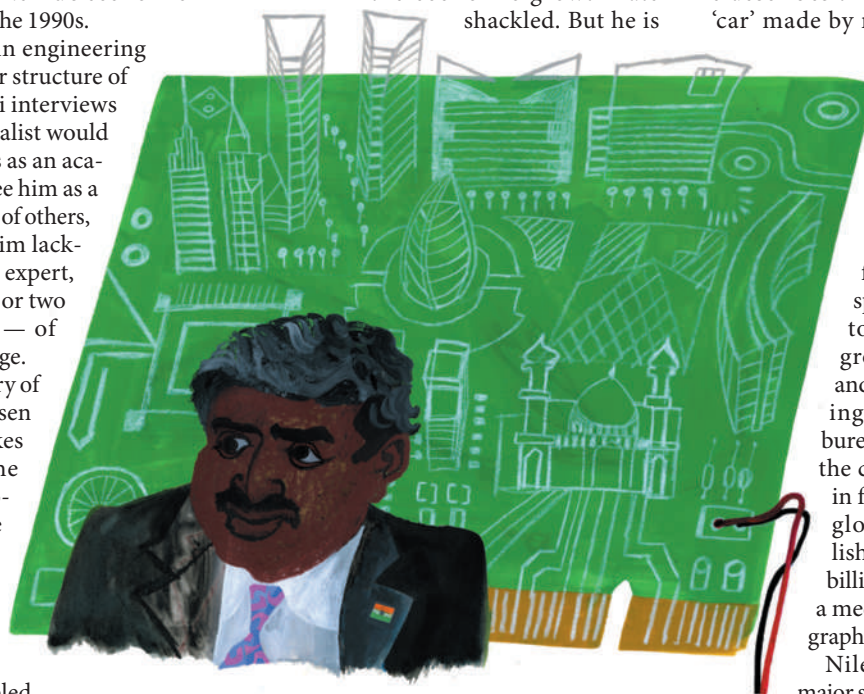
Nilekani pleads for innovation in business and governance. He refers to projects that enable slum children and illiterate farmers to gain access to computers and the Internet, and discusses grass-roots innovations based on the Indian tendency to improvise. He describes the *jugaad*, a multipurpose 'car' made by rural mechanics who put together whatever parts they can lay their hands on. Regrettably, he does not devote much space to India's strengths in advanced technology development.

Nilekani sees a great future for India, provided it speeds up economic reforms towards more market-led growth in the global arena, and adopts new ways of thinking to counter statism and bureaucracy. He is heartened by the change in people's attitude in favour of entrepreneurship, globalization and the English language. He hails India's billion-strong population, with a median age of 23, as its demographic dividend.

Nilekani notes that "all our major strengths have come together and matured at the same time". By using the rare opportunity provided by a constellation of domestic and international factors, he believes that India can reach farther and higher. Yet, he warns, such windows of opportunity do not remain open for long.

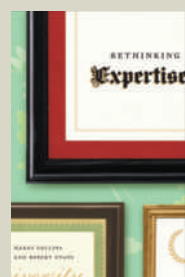
Hope has been an enduring characteristic of the Indian civilization. At a time when possibility has just encountered feasibility, *Imagining India* carries a strong message of hope. ■

L. K. Sharma is former foreign correspondent of *The Times of India* and editor of the book *Innovative India Rises*.
e-mail: sharmalk@gmail.com



not handcuffed to dogma. Nor does he spare private enterprise from criticism. He agrees that it was the legacy of those socialist policies that brought the country certain advantages when it decided to take the path of economic liberalization in the 1990s. His own business has drawn on the pool of talent directly nurtured by the state.

Part of the Nehruvian legacy is the country's competence in science and technology, most of it built up in the public sector. This competence came in handy when the private sector became mature enough to use it in industries such as biotechnology, IT and electronics. The



Rethinking Expertise

by Harry Collins and Robert Evans
(Univ. Chicago Press, £15.50)

By classifying different types of expertise, Harry Collins and Robert Evans make the case for a radical rethink of what constitutes expertise and how to exploit it. "Collins and Evans put their points vividly, with elegant language and diagrams," wrote Robert P. Crease (*Nature* **450**, 350–351; 2007).



A Cultural History of Modern Science in China

by Benjamin A. Elman
(Harvard Univ. Press, \$17.95)

In this concise, accessible, but comprehensive book, Benjamin Elman describes the effects on science of the Jesuit mission in imperial China in 1600–1800, and the later influence of Protestants in the nineteenth century. By doing so, he places the emergence of modern science in China in historical context.

An eye on the Universe

Einstein's Telescope: The Hunt for Dark Matter and Dark Energy in the Universe

by Evalyn Gates

W. W. Norton: 2009. 320 pp. £18.99, \$25.95

Gravitational lenses — heavenly bodies that bend light rays by their gravity — have emerged as powerful tools in the astronomer's repertoire. Yet their usefulness was difficult to foresee. In the 1930s, Albert Einstein wrote sceptically that "there is no great chance of observing this phenomenon". Astronomer Henry Norris Russell titled them 'Perfect Tests of General Relativity that are Unavailable'. Only the optimistic Fritz Zwicky predicted that "the probability that ... gravitational lenses will be found becomes practically a certainty", although he did not live to see it.

Einstein's Telescope opens up this new view of the Universe. Originally proposed as a semi-crazy thought experiment by Johann Soldner in 1801, gravitational lensing was ignored for more than a century until Einstein addressed it in a paper in 1911. Gravitational light deflection made headlines when astronomers Arthur Eddington and Frank Dyson measured offsets in the positions of stars viewed near the Sun's edge during the solar eclipse of 29 May 1919, confirming a prediction of Einstein's general theory of relativity. But it was another 60 years before the first double image of a lensed quasar was discovered in 1979, and the technique's use in cosmology began.

Conceptually simple, gravitational lensing occurs when a light ray is attracted by a cosmic body, its path becoming bent as if being focused by a lens. The amount by which the ray is deflected is proportional to the mass of the 'lens', and inversely proportional to the closest distance by which the ray passes the lens. Lensing is seen in astrophysical systems over a huge range of scales: from nearby stars and planets to distant and massive clusters of galaxies. And

it can be applied to many problems, from the detection of invisible dark matter and the magnification of very distant galaxies to the search for Earth-like planets around other stars. Given the method's power, one wonders why it has taken so long for a book to be written about it.

Author Evalyn Gates is an active scientist with a background in particle physics, cosmology and public outreach. This combination gives her the ideal tools to describe gravitational lensing. She addresses the book to the scientifically curious layperson — and delivers. Gates does an excellent job of directing readers through diverse astrophysical scenarios, including the 'strong' and 'weak' variants of lensing. She comprehensibly describes observations such as the distortion of background galaxies seen through a giant galaxy cluster and the surge in the brightness of stars caused by compact lenses moving in front of them.

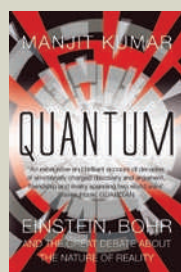
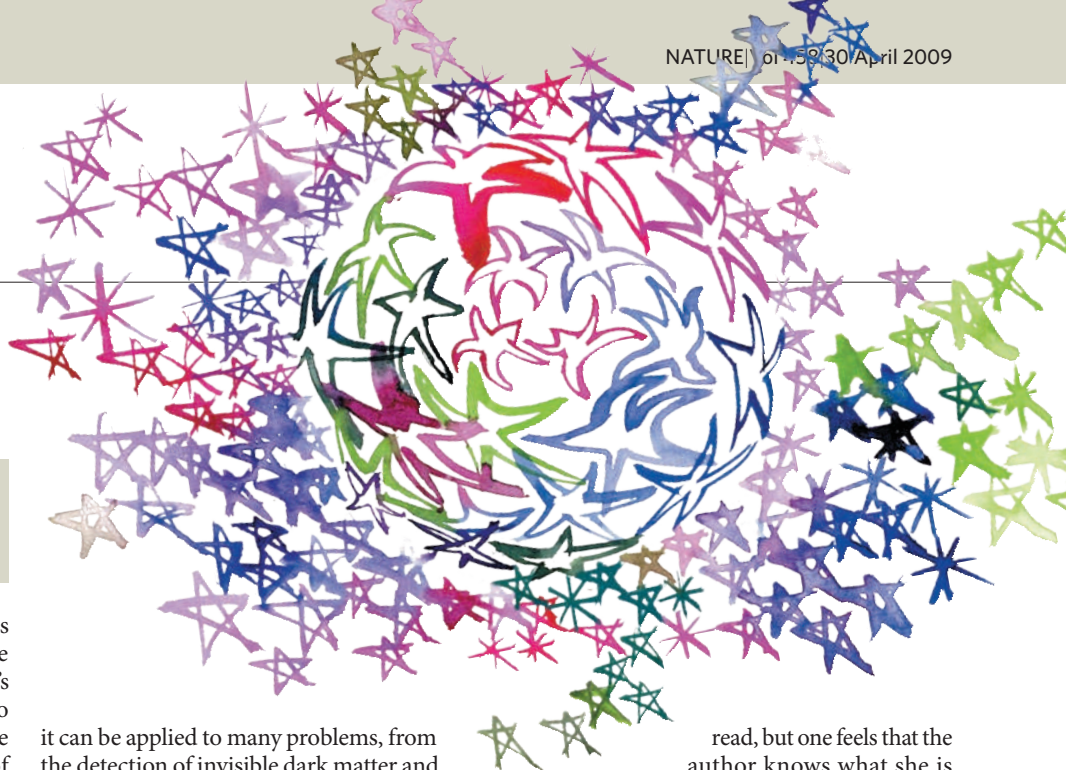
Gates uses the search for dark matter and dark energy — the two greatest mysteries in modern physics and astronomy — as her main driver. With this in mind, she explores fields from galaxy clusters to the searches for weakly interacting particles and from distant supernovae to the cosmic microwave background. But she also touches on many other issues, such as star and planet formation and galaxy evolution. She wants the reader to understand the underlying physics questions. Some of these are complex, so the book is not always an easy

read, but one feels that the author knows what she is writing about. Guiding the reader through multiverses, accelerated cosmic expansion or the physics of exploding stars, Gates clearly explains the problems and challenges that physicists face today. She uses apt analogies and original comparisons, many of which I had not heard before. They are so appropriately chosen that I found myself eagerly awaiting the next good metaphor.

Gates discusses the major challenges in cosmology. Not all of these questions will be solved with gravitational lensing, but in many cases it will contribute to the answer. Some of the topics appear to be slightly out of focus; but then, we learn, a gravitational lens does not have a single focal point.

Einstein's Telescope is comprehensive, including a glossary and a good index, referencing original literature and websites as well as offering detailed explanations of physical mechanisms. It provides food for thought for both expert and general readers. Through Gates's ambitious book, everyone will appreciate the puzzles of the dark Universe — and the power and beauty of gravitational lensing. ■

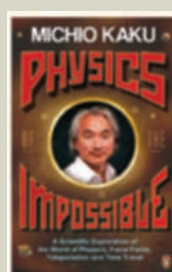
Joachim Wambsgans is professor of astronomy at Zentrum für Astronomie, Universität Heidelberg, Mönchhofstrasse 12-14, 69120 Heidelberg, Germany. He is currently on sabbatical as Bohdan Paczyński Visitor at Princeton University, New Jersey 08544, USA. e-mail: jkw@uni-hd.de



Quantum: Einstein, Bohr, and the Great Debate about the Nature of Reality

by Manjit Kumar (Icon Books, £9.99)

In this popular history of quantum mechanics, Manjit Kumar focuses "on the long-running debate between Niels Bohr and Albert Einstein, which took place from the mid-1920s through to the mid-1950s, over the adequacy of the quantum theory as a framework for fundamental physics", explained reviewer Don Howard (*Nature* **456**, 706-707; 2008).



Physics of the Impossible: A Scientific Exploration of the World of Phasers, Force Fields, Teleportation and Time Travel

by Michio Kaku (Penguin, £9.99)

Theoretical physicist Michio Kaku looks at various 'impossibilities' drawn from science fiction. He asks how new technology might help us achieve currently impossible phenomena, such as invisibility, and explains why others, such as precognition, would break the laws of physics.

CLIMATE CHANGE

Too much of a bad thing

Gavin Schmidt and David Archer

There are various — and confusing — targets to limit global warming due to emissions of greenhouse gases. Estimates based on the total slug of carbon emitted are possibly the most robust, and are worrisome.



It is one thing to agree on a goal of international policy, quite another to achieve it. The 192 signatories of the 1992 United Nations Framework Convention on Climate

Change (including the United States) have committed themselves to reducing the emissions of carbon dioxide and other greenhouse gases to avoid “dangerous interference in the climate system”. But policy-makers around the world are still trying to figure out how, specifically, to do that.

The European Union has adopted a goal of keeping temperatures below 2°C above pre-industrial levels. Others argue for a stabilization of atmospheric CO₂ concentrations at 350 parts per million (p.p.m.), or 450 p.p.m., or higher. In the United States, the administration of President Barack Obama has proposed reducing emissions by 80% by the year 2050. These schemes are all intended to be solutions to the same problem. But relating and comparing one to another is not straightforward. For instance, solutions to what level of atmospheric CO₂ is required to avoid a 2°C temperature rise, and what emission pathways might achieve that goal, are still unclear. Papers elsewhere in this issue by Meinshausen *et al.*¹ and Allen *et al.*² explore the uncertain relationships between carbon emissions and climate response, with the aim of better estimating how much additional CO₂ might indeed be too much.

Meinshausen and colleagues¹ (page 1158) take a comprehensive probabilistic approach, combining the uncertainties in climate sensitivity and carbon-cycle feedbacks, and integrating the two over a large range of potential emission pathways. Their target is to avoid a peak global mean warming from the pre-industrial level of more than 2°C (equivalent to a further rise of about 1.2°C from today). We must note here that there is nothing special about 2°C that would make warming of less than this magnitude ‘safe’. It is more analogous to a speed limit on a road, and is a guide to the scale of the problem. With 2°C of global warming (more over land and at the high latitudes), Earth would probably be warmer than it had been in millions of years — a huge change.

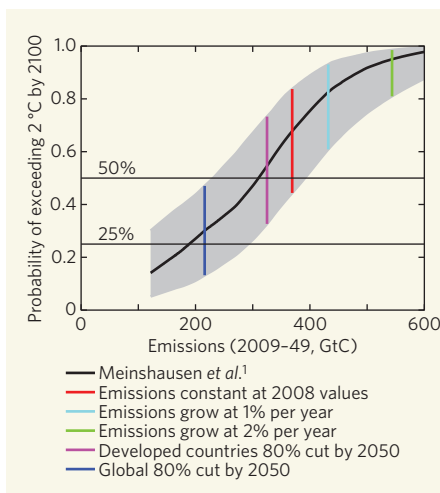


Figure 1 | The 2°C lottery. The black line shows the probability of the peak global mean temperature exceeding 2°C above pre-industrial levels before the year 2100 as a function of the integrated emissions from 2009 to 2049. The graph is adapted from the base case of Meinshausen *et al.*¹, including uncertainty ranges. Also shown are the cumulative emissions under various scenarios. Red, emissions constant at 2008 values until 2050. Light blue, growth in emissions continues at 1% per year until 2050 and then falls rapidly. Green, growth in emissions continues at 2% per year until 2050 and then falls rapidly. Purple, an 80% cut in emissions by 2050 (linearly applied, starting in 2010) from developed countries only, while developing-country emissions continue to grow at 1% per year. Dark blue, an 80% cut in emissions by 2050 from all countries.

Meinshausen *et al.* find that the maximum temperature that Earth will experience to the year 2100 depends most reliably on the total amount of CO₂ emitted to the year 2050, rather than on the final stabilized CO₂ concentration. Their base-case estimate is that the total emissions from today (2009) to 2050 need to stay below 190 GtC (equivalent to 700 GtCO₂; 1 GtC = 10¹² kg of carbon) for us to have a good chance (75%) of staying below 2°C (Fig. 1). The probability drops below 50% if we emit more than 310 GtC in that time. This is significantly less than the amount of carbon contained in proven reserves of gas, oil and coal,

let alone reserves of non-traditional fossil-fuel sources such as tar shales, oil sands or methane hydrates. Last year, we probably emitted more than 9 GtC, and this has been increasing at around 1–3% a year. At that rate, we will reach 190 GtC in under 20 years.

Allen and colleagues² (page 1163) take a slightly different tack, using a combined climate and carbon-cycle model, and varying uncertain parameters, to produce a series of simulations that attempt to span the range of projections that are consistent with already observed changes. They agree with Meinshausen *et al.* that it's the total slug of carbon that matters most, and define a term they call the cumulative warming commitment (CWC) as the peak temperature change expected as a function of the total anthropogenic carbon.

Comparing the bottom-line results from the two studies is tricky because of the use of different units, different base periods and different experimental design³. However, given that humans have already emitted roughly 520 GtC to the end of 2008, Allen and colleagues' best-estimate CWC — 2°C per 1,000 GtC emitted from 1750 to 2500 (compared with 2000–2050 in Meinshausen *et al.*) — implies that another 480 GtC would put us over 2°C with more than 50% likelihood. This is broadly consistent with the 310-GtC estimate from Meinshausen *et al.* over a much shorter time frame. For comparison, two scenarios with cuts of 80% in emissions by 2050, in developed countries and globally, give an additional 325 GtC and 216 GtC, respectively (Fig. 1).

A lot rides on the questions addressed in these papers, and they are unlikely to be the last words written on the topic. There is certainly room for further debate on the definition of ‘dangerous’; the maximum global temperature is a good place to start, but ice sheets and sea level, for example, probably depend on the integrated climate impact rather than on peak warming^{4–6}. Also, these studies^{1,2} use the traditional, short-term ‘Charney’ climate sensitivity, which includes some fast feedbacks such as variations in atmospheric water vapour and clouds, but not the slower feedbacks such as changes in vegetation or ice sheets, or feedbacks in atmospheric aerosols. The true

sensitivity of the Earth system may well be higher⁷, implying that any temperature-based target will become progressively harder to maintain as slower feedbacks kick in.

Finally, both studies^{1,2} make different assumptions about how non-CO₂ factors (anthropogenic methane, ozone, black carbon, sulphates and so on) will change. These effects can't be shoehorned into the same cumulative-emissions metric as CO₂ because their effects over time are much more closely tied to contemporaneous emission levels. However, they remain a tempting additional policy target that might usefully limit near-term temperature rises⁸.

The bottom line? Dangerous change, even loosely defined, is going to be hard to avoid. Unless emissions begin to decline very soon, severe disruption to the climate system will entail expensive adaptation measures and may eventually require cleaning up the mess by actively removing CO₂ from the atmosphere. Like an oil spill or groundwater contamination,

it will probably be cheaper in the long run to avoid making the mess in the first place. ■

Gavin Schmidt is at the NASA Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA. David Archer is in the Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA. e-mails: gschmidt@giss.nasa.gov; d-archer@uchicago.edu

1. Meinshausen, M. *et al.* *Nature* **458**, 1158–1162 (2009).
2. Allen, M. R. *et al.* *Nature* **458**, 1163–1166 (2009).
3. Allen, M. R. *et al.* *Nature Reports Climate Change* www.nature.com/climate/2009/0905/full/climate.2009.38.html (2009).
4. Hansen, J. *et al.* *Atmos. Chem. Phys.* **7**, 2287–2312 (2007).
5. Mann, M. E. *Proc. Natl Acad. Sci. USA* **106**, 4065–4066 (2009).
6. Smith, J. B. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 4133–4137 (2009).
7. Hansen, J. *et al.* *Open Atmos. Sci. J.* **2**, 217–231 (2008).
8. Shindell, D. & Faluvegi, G. *Nature Geosci.* **2**, 294–300 (2009).

See also Editorial, page 1077, and online at www.nature.com/climatecrunch

CELL BIOLOGY

Another way to get rid of fat

Rudolf Zechner and Frank Madeo

When starved, cells resort to breaking down their assets — proteins, lipids and even whole organelles. An investigation of lipid metabolism indicates that one process — autophagy — targets all three cellular components.

Fatty acids are essential to all organisms — as substrates for energy production, as precursors of membrane lipids and as signalling molecules that control various cellular processes, including gene expression. They are stored as triglycerides in highly dynamic organelles called lipid droplets¹ and, when necessary, are re-released by the process of lipolysis. Established models of lipid storage and breakdown have undergone substantial revision in recent years. The latest adjustment is offered by Singh *et al.*² (page 1131 of this issue): their spectacular findings suggest that autophagy, the pathway by which excess or damaged organelles and proteins are degraded³, also mediates fat mobilization and breakdown in liver cells (hepatocytes).

A rapid flux of fatty acids into and out of lipid droplets occurs through the deposition and degradation of triglycerides in adipose tissue and in other tissues (liver, heart, muscle and testis) and cells (macrophages) that require abundant fatty acids. Hydrolytic enzymes called lipases mediate lipid breakdown (catabolism), and for 40 years it was believed that hormone-sensitive lipase (HSL) acted alone in the catabolism of triglycerides. But the discovery of other essential lipases, such as ATGL, and regulatory proteins, including perilipin and CGI-58, indicated that the lipolytic pathway is much more complex than that⁴.

In fact, key aspects of lipid turnover remain

unclear. One mystery is the rapid turnover of triglycerides and cholesteryl esters (another component of lipid droplets) in hepatocytes, despite the cells' low concentrations of HSL and ATGL. By presenting compelling evidence for an autophagic mechanism mediating fasting-induced lipolysis in both mouse liver and culture-grown hepatocytes, Singh *et al.* resolve some of these issues.

In autophagy, cytoplasmic components and cellular organelles destined for degradation become trapped in double-membrane-bound vesicles called autophagosomes, and are then broken down in lysosomes with which the autophagosomes fuse³. This sequestration and lysosomal breakdown of autophagosomal contents is generally referred to as macroautophagy.

In a functional analogy to macroautophagy, Singh *et al.*² show that, under fasting conditions, the cytoplasmic protein LC3 and several other autophagy-related proteins are recruited to lipid droplets, where they form a double membrane that encloses droplet parts. These lipid-containing vesicles, termed autolipophagosomes, subsequently fuse with lysosomes, and their contents are degraded (Fig. 1).

The authors also show that the efficiency of this process of 'macrolipophagy' varies with the nutritional status of the mice. Feeding the

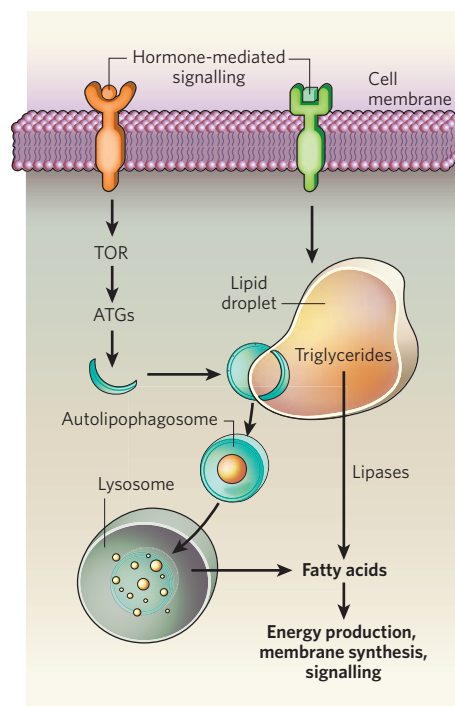


Figure 1 | Lipolysis and autophagy. Activation of intracellular signalling pathways mediated by the hormones glucagon and catecholamines can trigger the hydrolysis of triglycerides to fatty acids in lipid droplets by the process of lipolysis, which is mediated by lipase enzymes. Singh and co-authors² demonstrate that, during fasting, one such pathway in the liver involves TOR kinase and ATG proteins, and results in fatty-acid generation by an alternative mechanism. In a process they call macrolipophagy, portions of lipid droplets, or even whole droplets, become trapped inside the double-membrane-bound autolipophagosome vesicles and are transported to lysosomes, where they are degraded to fatty acids. Both lipolysis and macrolipophagy are inhibited by the hormone insulin.

animals a high-fat diet for an extended period (16 weeks) impairs autophagy-mediated breakdown of lipid stores in the liver, inducing a vicious circle in which — as the authors propose — increased fat ingestion may be associated with decreased fat removal and excessive lipid deposition in the liver. A similar response in obese people would explain how they might develop fatty-liver disease. Additionally, autophagy diminishes with age, possibly explaining the age-related deposition of fat in tissues where it does not belong. Singh and colleagues' results therefore provoke speculation that inducing autophagy through drug-mediated inhibition of TOR kinase, its master regulator protein, might ameliorate diet- or age-induced fatty-liver disease in humans. Intriguingly, resveratrol — a potent inducer of autophagy found naturally in red wine⁵ — prolongs the lifespan of mice fed a high-calorie diet⁶.

It remains to be seen how extensively macrolipophagy contributes to lipid homeostasis under physiological conditions, and whether it

is involved in fat mobilization in adipose tissue. Another question relates to the role of macrophages. Accumulation of cholesterol and cholesteryl esters in these cells leads to the formation of foam cells — an early indicator of atherosclerosis. Autophagy-facilitated hydrolysis of cholesteryl esters might mobilize cholesterol from lipid droplets and initiate its transport to the liver, a process called reverse cholesterol transport.

Moreover, the effect of autophagy on cellular lipid homeostasis is probably complex, and may not be restricted to macrolipophagy, because indirect contributions from the autophagy of cellular organelles such as mitochondria and peroxisomes are certainly possible. In their experiments, Singh *et al.*² deleted the gene encoding ATG7, a protein that probably affects all forms of macroautophagy. So the effects they observed on lipid homeostasis in the liver of mice lacking ATG7 may be multifactorial and not necessarily restricted to defective macrolipophagy alone. Accordingly, it is essential to determine the contribution of different subtypes of autophagy to lipid and energy homeostasis.

The exact mechanism of autolipophagosome assembly and the vesicles' subsequent fate also require detailed investigation. For instance, lipolytic hormones such as catecholamines cause lipid droplets to break down into smaller particles, streamlining fat catabolism in adipose tissue⁷. Is autolipophagosome formation the mechanism by which lipid-droplet fragmentation occurs?

Another question relates to the lipases that mediate autolipophagosome degradation. Lysosomal acid lipase is the only triglyceride and cholesteryl ester hydrolase enzyme known to exist in lysosomes and is therefore predicted to degrade these lipid ester components of autolipophagosomes. This enzyme also catabolizes the lipid components of internalized lipoproteins. Deficiency of lysosomal acid lipase causes Wolman's disease, a disorder characterized by the accumulation of triglycerides and cholesteryl esters in various tissues, and resulting in premature death. Excessive lipid storage in these patients might thus arise from defective breakdown of both external lipoproteins and lipid droplets in lysosomes. Notably, lipid-containing cytoplasmic particles have been observed in hepatocytes of both a rat model of Wolman's disease⁸ and patients with liver disease⁹; these particles could be unprocessed autolipophagosomes.

Finally, the exact role of LC3 must be established. Whereas Singh *et al.* show that this protein is essential for the formation of autolipophagosomes and for lipolysis, another study¹⁰ has found that it is involved in lipid and lipid-droplet formation.

Macrolipophagy adds another level of complexity to the breakdown of fat during fasting. In addition to lipase-mediated lipolysis, this process offers an alternative target

for the control of fat breakdown in hepatocytes and possibly other cells and organs. Beside the potential for drug intervention, it is worth remembering that an old adage of innumerable human cultures holds that fasting offers the best protection from various diseases, including metabolic disorders. Macrolipophagy might be a crucial component of this widespread practice. ■

Rudolf Zechner and Frank Madeo are at the Institute of Molecular Biosciences, University of Graz, Graz 8010, Austria.

e-mail: rudolf.zechner@uni-graz.at

1. Martin, S. & Parton, R. G. *Nature Rev. Mol. Cell Biol.* **7**, 373–378 (2006).
2. Singh, R. *et al.* *Nature* **458**, 1131–1135 (2009).
3. Levine, B. & Kroemer, G. *Cell* **132**, 27–42 (2008).
4. Zechner, R. *et al.* *J. Lipid Res.* **50**, 3–21 (2009).
5. Oipari, A. W. Jr *et al.* *Cancer Res.* **64**, 696–703 (2004).
6. Baur, J. A. *et al.* *Nature* **444**, 337–342 (2006).
7. Brasaemle, D. L. *J. Lipid Res.* **48**, 2547–2559 (2007).
8. Kuriwaki, K. & Yoshida, H. *Pathol. Int.* **49**, 291–297 (1999).
9. Hayashi, H., Sameshima, Y., Lee, M., Hotta, Y. & Kosaka, T. *Hepatology* **3**, 221–225 (1983).
10. Shibata, M. *et al.* *Biochem. Biophys. Res. Commun.* **382**, 419–423 (2009).

X-RAY ASTRONOMY

When appearances are deceptive

Michael Shull

The sharpest X-ray image ever obtained of a portion of the Milky Way resolves a seemingly diffuse X-ray emission into discrete sources. These sources are likely to be stars of the 'garden variety' in the Sun's vicinity.

With blurry vision, one can imagine strange and wonderful things that later turn out to be erroneous. Astronomers have learned this lesson before, from the hypothetical canals on Mars¹ to claims of rotating spiral arms in the Andromeda galaxy². Even the imaginary 'man in the moon' figure may be the result of poor vision.

In its early years, X-ray astronomy had similar difficulties interpreting the diffuse, extragalactic X-ray background (XRB) emission from the sky³. Was this X-ray glow produced by hot gas spread between the galaxies or a blur of unresolved point sources? Only with the

development of modern X-ray detectors and telescopes, particularly the Chandra X-ray Observatory⁴, was most of this cosmic XRB radiation resolved into discrete sources. The 'background' turned out to be a myriad of distant quasars and other galaxies whose bright cores are powered by supermassive black holes⁵.

On page 1142 of this issue, Revnivtsev *et al.*⁶ address a similar problem — that of deciphering the hitherto unresolved X-ray emission from the Milky Way's disk of gas and stars, the Galactic X-ray ridge emission^{7,8}. Their analysis resolves most of this apparently diffuse X-ray emission into discrete sources. The results indicate that many of these sources are stars, one of the less exotic early suggestions^{9,10}.

As its name suggests, the Galactic X-ray ridge is a ridge of X-ray emission in the Galaxy: at energies above several kiloelectronvolts (keV), the diffuse XRB emission exhibits an excess near the Galactic disk that extends a few hundred parsecs (about 800 light years) in height above the disk.

Its spectrum⁹ is characteristic of a hot plasma at 100 million kelvin, including a prominent emission line of iron at 6.7 keV. Such emission is similar to that from hot intergalactic gas confined to clusters of galaxies. The problem with this hot-plasma description of the Galactic ridge emission is that the radiative energy losses from hot electrons in the plasma would be enormous¹¹, perhaps 10^{43} erg s⁻¹, exceeding all likely sources of heat in the Galaxy, even supernovae. Furthermore, the gravity in the Milky Way disk is much too weak to confine this hot plasma, which should instead be flowing away at large velocities in a Galactic wind. This phenomenon has been observed in other galaxies, such as M82 (Fig. 1).

The sharp optics of the Chandra Observatory allowed Revnivtsev *et al.*⁶ to detect point



Figure 1 | Galactic wind. A horizontal disk of stars (white) and a perpendicular galactic wind of hot plasma (purple) are visible in this image of the starburst galaxy M82. Revnivtsev and colleagues⁶ show that the apparently diffuse X-ray emission from the Milky Way's disk is not due to a Galactic wind but instead to discrete sources such as stars. This colour-coded, visible-light image combines observations taken by the WYIN 3.5-metre telescope and the Hubble Space Telescope.

NOAO/AURA/NSF/WYIN CONSORTIUM

sources in an ultra-deep image obtained from a million-second (equivalent to about 12 days) exposure of a small (16×16 arcminute square) region of sky near the Galactic Centre (see Fig. 1 on page 1143). In total, they detected 473 X-ray sources with a reliable degree of confidence. These sources account for a substantial fraction of the X-ray emission: in the energy range 6.5–7.1 keV, and particularly near the prominent 6.7-keV iron emission line, the authors' analysis resolves more than 80% of the emission into discrete sources, most of which are probably stars.

Revnivtsev and colleagues' work demonstrates the power of combining ultra-long X-ray sky exposures with exquisite telescope optics. In many ways, the Chandra Observatory marks a high point of X-ray imaging, with mirrors capable of resolving objects to an accuracy of 0.5 arcsecond — the size of a small coin viewed from a distance of 10 kilometres. Most of the X-ray sources found in the Chandra image are probably around the same distance away as the Galactic Centre is from Earth, which is approximately 8.5 kiloparsecs (27,700 light years). At this distance, the brightest X-ray sources have luminosities of 10^{32} erg s⁻¹, just 3% of the Sun's total luminosity at all wavelengths.

The minimum luminosities detectable, with the sensitivity of the mirrors and cameras aboard Chandra, are 100 times fainter, around 10^{30} erg s⁻¹. These X-ray sources are likely to be white dwarfs with luminosities of 10^{31} – 10^{32} erg s⁻¹ accreting matter from companion stars, and binary stars with strong coronal activity — akin to flares on the Sun, but with X-ray luminosities of 10^{31} erg s⁻¹ or less. Such stellar X-ray sources are of the common 'garden variety' in the Sun's neighbourhood¹². However, at the distance of the Galactic ridge from Earth, their combined light becomes a diffuse blur, the X-ray equivalent of the many stars that make up the Milky Way, as Galileo first saw with his telescope in visible light.

Regarding the value of high-resolution imaging, it is worth reflecting on the greater significance of Revnivtsev and colleagues' observations. The resolution of the seemingly diffuse X-ray emission into discrete sources eliminates a physically implausible explanation of that emission that required enormous energy sources of hot plasma. The discrete-source model offers the chance to apply techniques of optical (visible-light) astronomy, such as stellar-population synthesis and surface-brightness fluctuations — both based on the assumption that galaxies are made up of a finite number of stars — to study galaxies other than our own in the X-ray waveband. For example, the apparently diffuse X-ray emission of distant galaxies may also be dominated by unresolved faint stellar sources and supernova remnants^{7,9,10}.

As with the extragalactic X-ray background, the sources that produce the Galactic ridge X-ray emission proved to be less exotic than some early suggestions. Perhaps there are lessons here, and astronomers should bear

that in mind as they explore sources of faint backgrounds at many wavelengths, from X-ray to infrared. As Revnivtsev and colleagues' work⁶ demonstrates, sometimes the exotic explanation can be set aside by more accurate imaging and spectroscopy.

Michael Shull is in the Department of Astrophysical and Planetary Sciences, University of Colorado at Boulder, Colorado 80309, USA. e-mail: michael.shull@colorado.edu

1. Lowell, P. *Astrophys. J.* **26**, 131–140 (1907).
2. van Maanen, A. *Astrophys. J.* **81**, 336–337 (1935).
3. Giacconi, R., Gursky, H., Paolini, F. & Rossi, B.

Phys. Rev. Lett. **9**, 439–443 (1962).

4. Weisskopf, M. C. *et al. Astron. Soc. Pacif.* **114**, 1–24 (2002).
5. Hasinger, G. *et al. Astron. Astrophys.* **365**, L45–L50 (2001).
6. Revnivtsev, M. *et al. Nature* **458**, 1142–1144 (2009).
7. Worrall, D. M., Marshall, F. E., Boldt, E. A. & Swank, J. H. *Astrophys. J.* **255**, 111–121 (1982).
8. Warwick, R. S., Pye, J. P. & Fabian, A. C. *Mon. Not. R. Astron. Soc.* **190**, 243–260 (1980).
9. Koyama, K., Makishima, K., Tanaka, Y. & Tsunemi, H. *Publ. Astron. Soc. Jpn* **38**, 121–131 (1986).
10. Worrall, D. M. & Marshall, F. E. *Astrophys. J.* **267**, 691–697 (1983).
11. Tanaka, Y. *Astron. Astrophys.* **382**, 1052–1060 (2002).
12. Sazonov, S., Revnivtsev, M., Gilfanov, M., Churazov, E. & Sunyaev, R. *Astron. Astrophys.* **450**, 117–128 (2006).

MINIATURE DEVICES

Voyage of the microrobots

Metin Sitti

Nanobots — tiny robots that can be injected into the body to perform medical procedures — are the stuff of science fiction. Swimming microrobots propelled by artificial flagella bring that fantasy closer to reality.

In the classic film *Fantastic Voyage*, a miniaturized submarine and its crew are injected into a coma victim in a perilous mission to destroy the blood clot that threatens the patient's life. Far-fetched as this might seem, tiny, tetherless robots — albeit without a human crew — might one day be able to access small spaces inside the human body that can currently be reached only using invasive surgical methods. This could revolutionize medicine, being potentially cheaper, less painful and more flexible than surgery, and enabling diseases to be precisely and locally targeted for diagnosis and treatment.

Of course, medical microrobots are a long way off, as there are many technical problems to overcome — such as how to make, power and steer them. But reporting in *Applied Physics Letters*, Zhang *et al.*¹ propose possible solutions for all three of these challenges. Building on previous studies^{2,3} of fluidic propulsion in microbotic design, they have made a swimming robot that is the size of a bacterium.

The authors were inspired by the flagella of bacteria, which enable the organisms to swim efficiently across micrometre distances. Their microrobot¹ thus consists of an artificial, helical flagellum attached to a magnetic 'head'. The flagellum turns when placed in a weak, oscillating magnetic field created by three orthogonal pairs of electromagnetic coils. The spinning flagellum thus propels the robot efficiently through liquid at speeds of 1–2 micrometres per second. For comparison, typical bacteria can swim at up to 20–100 micrometres per second, depending on the type.

Zhang *et al.* produce their robots by etching them out of thin films of semiconductors that have been deposited on a substrate. The flagellum is designed so that it scrolls itself into a

helix on release from the substrate. Such fabrication methods are fine for simple robots, but medical devices will require on-board sensors, drug-delivery tools, computing components and so on. The microrobots of the future will thus need to be assembled using more complex methods.

By using magnetic actuation to power their microrobots, Zhang *et al.* avoid the need to place actuators or energy sources on the robots' bodies. Furthermore, external actuation should provide physicians with a direct way of controlling and steering robots inside patients, because the direction of the robots' movements will alter in response to changes in the direction of the magnetic field. However, magnetic forces decay exponentially with distance, and so very high magnetic-field gradients would be required to steer robots deep inside the body. In some cases, such gradients might heat and damage tissues. Another potential problem with the authors' actuation method is that it would require expensive and bulky magnetic instrumentation in the operating theatre.

Many groups are therefore investigating on-board actuation and powering methods for microrobots. A recent trend is to use living cells as on-board actuators that are fuelled by chemical energy — specifically, by ATP molecules, which fuel most biochemical reactions and are found throughout the body. Bacterial^{4,5} and algal⁶ cells have been integrated into micro- and nanofabricated robot bodies, and can propel the devices through still fluids. This is fine for regions of the body, such as the eye, that contain motionless fluids, but cell-based propulsion mechanisms wouldn't provide microrobots with enough force to travel through liquids flowing at high speed, such as blood. Such 'cyborg' microrobots could,

however, use chemical gradients, light or other stimuli to passively or actively control the motion of the cell component.

The advantages of cellular 'motors' are that the overall system can be compact, fast and inexpensive; mass production of the microrobots would also be cost-effective if self-assembly processes were used. Furthermore, the cellular propulsion or actuation component could be multifunctional, perhaps serving as a biochemical sensor or navigation controller. Such systems are, however, limited by the often stochastic nature of cellular motion, and by the relatively brief lifetimes of cells (depending on the environments in which the robots are used).

Future challenges include finding methods for detecting, localizing and communicating with microrobots inside the body, and making and integrating on-board diagnostic and therapeutic tools. The safety of patients during operations must also be guaranteed: for example, it remains to be shown that microrobots can operate in the dynamic environments of the human body in a reliable, robust manner. And ideally, the instruments required for use in microrobots should be cheap and portable.

Moreover, methods must be found to program and control armies of microrobots, as this will be necessary if such devices are to treat large areas of the body, to increase the speed and success of medical operations, and to deliver sufficient amounts of drugs to their targets. Zhang and colleagues' method¹ of external actuation could in principle, be used to steer and coordinate such armies. But although this approach would certainly allow fast and effective control, it would also require complex instrumentation, robot design and fabrication. Alternatively, passive steering could be used to control large numbers of cyborg microrobots, for example by genetically engineering attached cells to be attracted towards a specific biochemical signal. This approach would be energy efficient and would not require bulky external equipment.

With so many engineering and medical challenges to be solved, clinically usable microrobots may not be viable for decades to come. Nevertheless, the demonstration of methods for making, powering and steering micro-metre-scale robots^{1,4-6} are the first crucial steps towards such systems. ■

Metin Sitti is in the Department of Mechanical Engineering, Carnegie Mellon University, 320 Scaife Hall, Pittsburgh, Pennsylvania 15213-3890, USA.
e-mail: sitti@cmu.edu

ECOLOGY

Speciation affects ecosystems

Ole Seehausen

Evidence that speciation and adaptive radiation can change the properties of an ecosystem is a reminder of the pressing need to integrate ecosystems science and evolutionary biology.

Differences between ecosystems affect the rate and direction of evolutionary diversification. But does evolutionary diversification affect the functioning of ecosystems? On page 1167 of this issue, Harmon *et al.*¹ present experimental data from work with predatory stickleback fish that show that it does.

This insight comes at an opportune moment, as ecosystems ecology and evolutionary biology have existed separately for more than a century, and their integration is long overdue. Evolutionary biology is all about the dynamics that arise from interactions of genotypes with their environments and interactions among genotypes, which scale up to the dynamics of biodiversity in space and time. Ecosystems ecology, on the other hand, looks at dynamics predominantly in terms of fluxes of matter and energy, implicitly or explicitly assuming that these are unaffected by genotypic variation within species.

Building blocks for a synthesis have begun to emerge in several contexts, all emphasizing feedbacks between ecosystems and evolution². Besides the well-established fact that environments change the course of evolution, evolution also changes the environment, and in turn, the environment changes evolution further, and so forth. The effects of a population on its environment are determined by its genotypic composition and evolutionary history; these effects change the form and strength of selection on the population itself ('niche construction') and of selection on other species ('ecosystems engineering'), and so determine the course of further evolution.

This feedback loop occurs on all timescales. Organisms produce effects on the environment that can persist over geological time and modulate macroevolution (that is, the origins of new taxonomic groups and biological assemblages)³. And evolution within a species — microevolution — can change ecosystem function. For instance, the evolution of predator life-history over the course of decades and centuries can have a profound impact on lower trophic levels, thereby driving changes in nutrient flux and primary production (algal growth)⁴, and influencing the further course of evolution in the predator itself⁵. Evolution of anti-predator defence can be so fast that it can even shape cyclical predator-prey dynamics, thought of as classical ecological dynamics⁶. But what has remained unknown is whether the splitting of one species into two affects ecosystem function.

This is where Harmon and colleagues¹ come in. They show that stickleback adaptive radiation — whereby one species splits into two or more, while adapting to alternative ecological resources^{7,8} — has considerable effects on the composition and abundance of species at lower trophic levels. These consequences are then reflected in changes in the aquatic environment in terms of primary production, the composition of dissolved organic materials and light penetration — ecosystem features that in turn are likely to affect the course of adaptive radiation.

After the retreat of the glaciers at the end of the last ice age, marine threespine sticklebacks colonized fresh waters, and split into many nascent species adapted to different lifestyles and environments. In several lakes along the Pacific rim of British Columbia, Canada, two coexisting species evolved through 'ecological speciation'⁹, one adapted to life at the lake edge (the benthic species) and the other to life in the open water (the limnetic species). In other lakes, a single stickleback species is found. The solitary species tend to be ecological generalists, although their levels of specialization may vary between lakes.

Harmon *et al.*¹ used experimental mesocosms to address the question of whether it made a difference to ecosystem function that sticklebacks had specialized and speciated. They inoculated replicate mesocosms with the same assemblages of open-water and benthic invertebrate prey, but then included fish in four different ways: as a single generalist stickleback species; as each of two different specialized species; or as a sympatric (coexisting) pair of specialized species.

Relative to the unspecialized species, each specialized species on its own produced a reduction of primary production and an increase ('upregulation') of benthic invertebrates (Fig. 1). However, large benthic invertebrates were more strongly upregulated in the presence of a limnetic stickleback species, whereas small to large zooplankton (plankton consisting of tiny invertebrates) were upregulated in the presence of benthic sticklebacks. Even though these stickleback species were taken from the same lake, the implication is that adaptive differences between stickleback populations in different lakes may affect ecosystem structure and function. When both species of stickleback were present, only the smallest zooplankton and benthic invertebrates were upregulated. The implication here is that adaptive radiation within a lake changed

1. Zhang, L. *et al.* *Appl. Phys. Lett.* **94**, 064107 (2009).
2. Baudry, J. *et al.* *Nature* **437**, 862–865 (2005).
3. Edd, J., Payen, S., Stoller, M. L., Rubinsky, B. & Sitti, M. *Proc. Int. Conf. Intelligent Robots and Systems* 2583–2588 (IEEE/RSJ, 2003).
4. Darnton, N., Turner, L., Breuer, K. & Berg, H. C. *Biophys. J.* **86**, 1863–1870 (2004).
5. Behkam, B. & Sitti, M. *Appl. Phys. Lett.* **93**, 223901 (2008).
6. Weibel, D. B. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 11963–11967 (2005).

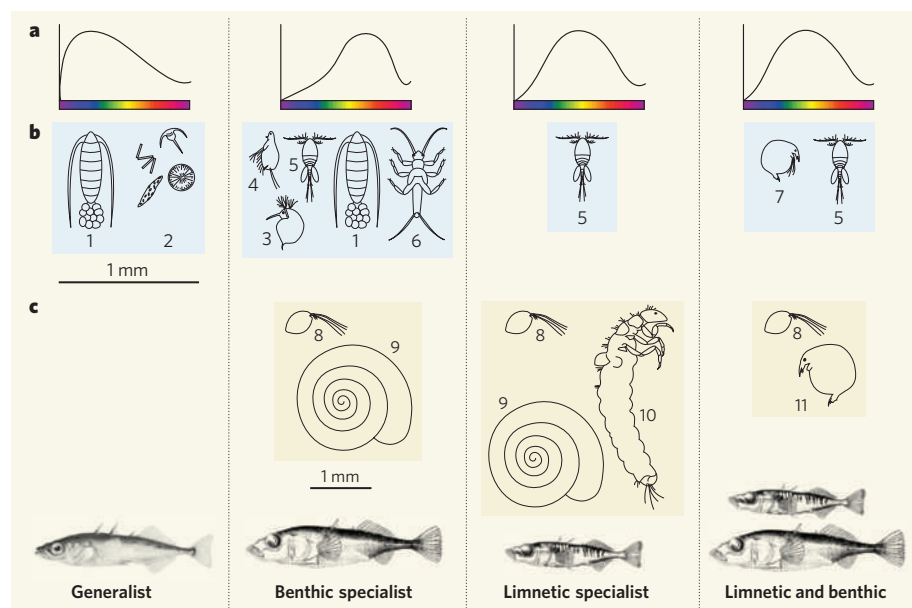


Figure 1 | Harmon and colleagues' experiments¹. Initially identical ecosystems in mesocosms diverged depending on the recent evolutionary history of the experimental population of three stickleback predator species in four combinations — ecologically generalized; lake edge (benthic); open water (limnetic); and limnetic and benthic together. All of these sticklebacks share a common ancestor that existed no more than 10,000 years ago, and probably much more recently. The implication is that rapid speciation may cause rapid changes in ecosystem structure and function. **a**, Idealized light spectra resulting from each treatment. **b**, Open-water prey that became relatively abundant. 1, *Skistodiaptomus* (copepod); 2, phytoplankton; 3, *Bosmina* (cladoceran); 4, *Diaphanosoma* (cladoceran); 5, *Cyclopoida* (copepod); 6, stonefly larva; 7, *Chydorus sphaericus* (cladoceran); 8, ostracod; 9, gastropod; 10, caddisfly larva; 11, benthic cladoceran. Note the different scales in **b** and **c**. (Fish images from refs 14, 15.)

its ecosystem from one dominated by phytoplankton (plankton consisting of microscopic plants) and large zooplankton to one dominated by small zooplankton and small benthic invertebrates. In mesocosms containing differently specialized species, or a species pair, the composition of dissolved organic material, and so the water's light-absorbance properties, also changed, altering the intensity and colour of light under water (Fig. 1a).

Intriguingly, the niche-construction effect of ecological speciation in predatory fish, through its effects on the light environment, may close a feedback loop to speciation. Variation in light exerts strong divergent selection between¹⁰ and within¹¹ lakes, because it influences vision and mating preference and so may determine whether or not speciation happens¹². Harmon *et al.* suggest that variation in the extent of niche construction between organisms might help to explain variation in the propensity to undergo adaptive radiation. If future research confirms this possibility, species diversity arising through adaptive radiation and ecosystem evolution would indeed seem to constitute a complete loop.

What next? Most obviously, there is a need to repeat the experiments with other, independently evolved pairs of stickleback species to see if the effects of adaptive radiation on an ecosystem are predictable. Then there is the question of whether such effects are indeed due to adaptive genetic divergence between

predator species, as Harmon *et al.* assume, or to non-genetic differences in behaviour or performance between sticklebacks collected from different habitats. Experiments with fish raised under identical laboratory conditions could be used to distinguish between these alternatives; eventually, this framework of research could be extended to adaptive radiations in which more

than two recently diverged species share the same ecosystem. Also, tests need to be devised to see if the effects seen in mesocosms apply in natural ecosystems. Altogether, these studies would permit the establishment of more general relationships between ecosystem functioning and adaptive radiation. As a by-product, such research will foster the synthesis of ecosystems science with evolutionary biology.

Implicit in this exciting news¹ is a warning relating to the rapid, human-induced collapse of species diversity in ecosystems dominated by recent adaptive radiations¹³. If the effects of adaptive radiation on the function of natural ecosystems are similar to those observed in the mesocosms, we should expect to see the collapse of nascent species diversity leading to severe perturbations in ecosystem functioning — a process that may in fact already be under way.

Ole Seehausen is at the Institute of Ecology and Evolution, University of Bern, and the Eawag Swiss Federal Institute of Aquatic Science and Technology, Center of Ecology, Evolution and Biogeochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland. e-mail: ole.seehausen@eawag.ch

1. Harmon, L. J. *et al.* *Nature* **458**, 1167–1170 (2009).
2. Post, D. M. & Palkovacs, E. P. *Phil. Trans. R. Soc. Lond. B* **364**, 1629–1640 (2009).
3. Erwin, D. H. *Trends Ecol. Evol.* **23**, 304–310 (2008).
4. Palkovacs, E. P. *et al.* *Phil. Trans. R. Soc. Lond. B* **364**, 1617–1628 (2009).
5. Palkovacs, E. P. & Post, D. M. *Evol. Ecol. Res.* **10**, 699–720 (2008).
6. Yoshida, T. *et al.* *Nature* **424**, 303–306 (2003).
7. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
8. Gavrillets, S. & Losos, J. B. *Science* **323**, 732–737 (2009).
9. Schluter, D. & McPhail, J. D. *Am. Nat.* **140**, 85–108 (1992).
10. Reimchen, T. E. *Evolution* **43**, 450–460 (1989).
11. Boughman, J. W. *Nature* **411**, 944–948 (2001).
12. Seehausen, O. *et al.* *Nature* **455**, 620–626 (2008).
13. Taylor, E. B. *et al.* *Mol. Ecol.* **15**, 343–355 (2006).
14. Kraft, C. E. *et al.* *Inland Fishes of New York (Online)* (Cornell Univ./NYSDEC, 2006).
15. McPhail, J. D. *Can. J. Zool.* **70**, 361–369 (1992).

SOLID-STATE PHYSICS

Lost magnetic moments

Richard Korytár and Nicolás Lorente

A neat study gives clear-cut evidence that when a wire made of a magnetic material such as iron is squashed to the atomic scale, the material's magnetism disappears via an exotic physical process.

Since the 1960s, it has been known that a magnetic impurity in a non-magnetic host metal is subject to a mechanism known as the Kondo effect. That is, below a characteristic temperature, the Kondo temperature, the metal's electrons screen the magnetic moment of the impurity^{1,2}. In this context, ferromagnetic metals, such as those commonly used to make magnets for holding notes on refrigerator doors, are intriguing. These materials, which are marked out by their ability to

retain magnetization, can in principle screen a magnetic impurity³. But can a ferromagnet screen one of its own atoms? On page 1150 of this issue, Calvo *et al.*⁴ show that this can be achieved when a ferromagnetic wire is reduced to atomic dimensions.

In their experiment, Calvo and colleagues created atomic-size ferromagnetic wires using two techniques known as scanning tunnelling microscopy and electromigration. These techniques allowed the authors to reduce the size

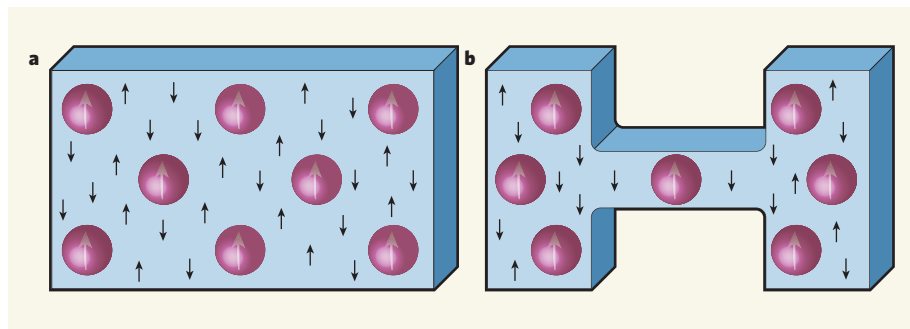


Figure 1 | When size matters in ferromagnetism. **a**, In a ferromagnetic wire, the total magnetic moment is due to the alignment of the individual magnetic moments of *d* electrons bound to the atomic sites of the material (white arrows). The *s* and *p* electrons, which act as if they are moving freely, make a negligible contribution to the magnetism of the material (black arrows). **b**, Calvo and colleagues⁴ show that, when a ferromagnetic wire is reduced to the size of a single atom, the interaction of the *d* electrons is decreased. In contrast, their interaction with the *sp* electrons is increased. The *sp* electrons efficiently screen the magnetic moment of the atomic contact, a manifestation of the Kondo effect.

of a junction between two metallic conductors made of pure ferromagnets — iron, cobalt or nickel — to the scale of a single atom, creating ferromagnetic, atomic-size contacts with full mechanical and thermal stability. Next, the authors measured how the conductance of the fabricated atomic contacts changed as a function of an applied voltage. For all three ferromagnets, the atomic contacts showed anomalies, in the form of peaks or dips, in their conductance at around zero volts. This is a sign of the Kondo effect, although not all zero-voltage conductance anomalies are indicative of it.

To confirm that the atomic contacts indeed showed Kondo behaviour, Calvo and co-workers studied the evolution of the conductance anomalies with temperature and found that it agreed with that expected for a system in the Kondo regime: the peaks and dips in the conductance curves disappeared with increasing temperature. What's more, they showed that the energy of the system, which is related to the temperature by the Boltzmann constant, followed a log-normal distribution — that is, the logarithm of the energy obeyed a Gaussian distribution. This observation is, again, in accord with the predictions of Kondo physics. Hence, the authors give clear proof that the Kondo effect is at work in these systems.

But their results are surprising. How can an atom that is a constituent member of a ferromagnet behave as a magnetic impurity screened by its identical neighbours? The Kondo effect usually requires two species of atom. Figure 1 shows a simple scheme for the screening process that takes place when a bulk ferromagnet is pulled into an atomic-size neck. In the bulk (Fig. 1a), all atoms contain electrons that have magnetic moments and are bound to the atomic sites; these '*d* electrons', so-called because they reside in the *d* atomic orbitals, where their magnetic moments tend to align, give rise to the total magnetic moment of the ferromagnet. In much smaller quantities, there are also '*sp* electrons', acting as if they are

moving freely, which occupy the *s* and *p* orbitals in the isolated atom. These electrons interact weakly with the localized *d* electrons. Electronic conduction mainly takes place through *sp* electrons.

But as Calvo *et al.*⁴ demonstrate, when the ferromagnet is constricted (Fig. 1b), the interaction between *d* electrons is greatly reduced. Moreover, the *sp* electrons increase their interaction with the *d* electrons because of the smaller number of neighbouring atoms, efficiently screening the magnetic moment

of the atomic contact. In other words, the magnetic moment of the atomic contact is effectively eliminated.

Calvo and colleagues' work shows that measurements of conductance can yield valuable information on the electronic properties of matter on the nanometre scale. Their results pave the way to studies of magnetism in atomic contacts and prompt further questions. Do non-magnetic atoms develop magnetism in non-magnetic materials as their dimensions are reduced? How does electronic transport take place? Does spin transport occur in addition to charge transport? We expect that, besides its fundamental implications for our understanding of how electrons interact in solid-state materials, this work will provide new insights into the properties of nanostructures and their use in fields such as magnetoelectronics, or indeed any technology operating on the scale where quantum mechanics rules.

Richard Korytar and Nicolás Lorente are at the Centre d'Investigació en Nanociència i Nanotecnologia CIN2 (CSIC - ICN), E-08193 Bellaterra, Spain.
e-mails: nicolas.lorente@cin2.es;
rkorytar@cin2.es

1. Yosida, K. *Theory of Magnetism* 155–304 (Springer, 1996).
2. Fulde, P. *Electron Correlations in Molecules and Solids* 3rd edn 290–295 (Springer, 1995).
3. Pasupathy, A. N. *et al. Science* **306**, 86–89 (2004).
4. Calvo, M. R. *et al. Nature* **458**, 1150–1153 (2009).

NEUROSCIENCE

A social hub for worms

Shawn R. Lockery

There are more connections in the human brain than there are stars in the Milky Way, so scientists use simple organisms to search for universal neural-circuit motifs. Their latest find is a neuron for social behaviour.

The nematode worm *Caenorhabditis elegans* has two distinct foraging styles: social and solitary¹. Social strains move rapidly on a patch of bacterial food, favour the borders of the patch (bordering), and cluster in tight little swarms when feeding (aggregation). Solitary strains, by contrast, move slowly on food, show no spatial preferences and feed alone. Whether or not a worm swarms might not seem a propitious place to dig for universal principles of neuronal function. On the contrary, however: by tracing the genetic basis of foraging style, neurobiologists have unearthed several unexpected findings^{2,3}. On page 1171 of this issue, Macosko *et al.*⁴ revisit social feeding and discover a 'hub-and-spoke' circuit for regulation of behaviour by a cluster of diverse sensory neurons.

Differences in *C. elegans* foraging styles can be explained by naturally occurring variations in the gene *npr-1*. This gene encodes the NPR-1

receptor, which resides on the surface of neurons and binds to protein fragments called neuropeptides, which are released as signals by other neurons. Solitary strains express a high-activity version of *npr-1*, whereas social feeders express a low-activity version¹; mutants with defective *npr-1* are likewise social. Nine types of neuron express *npr-1* (refs 4, 5), but, for technical reasons, it has been difficult to pinpoint which of these mediate social feeding.

Macosko *et al.*⁴ solve this problem by using, for the first time in *C. elegans*, the Cre-lox system — a powerful genetic technique for restricting the cellular expression pattern of distinct genes. They find that restoring the expression of normal NPR-1 protein specifically in a pair of neurons, known as RMG, converts *npr-1* mutants into solitary feeders. Importantly, killing these neurons in *npr-1* mutants also converts worms into solitary feeders, so RMG is probably turned off in solitary strains.

To determine the neural circuit within which RMG functions, the authors inspected the essentially complete anatomical wiring diagram of the *C. elegans* nervous system⁶. They find that this pair of neurons is connected by gap junctions — hollow proteins that allow electrical currents to flow between cells — to seven other neurons, six of which are sensory (Fig. 1a). This pattern of connectivity immediately suggested a hub-and-spoke arrangement with RMG at the centre. But how does this circuit function in social feeding?

In search of an answer, Macosko *et al.*⁴ chose the chemosensory neuron ASK, which integrates food and pheromone signals emitted by hermaphrodites⁷. In *npr-1*-defective social feeders, inactivating ASK and its sister chemosensory neurons reversed the effect of *npr-1* mutation, causing the animals to be solitary. Moreover, restoring the function of ASK, and just one other sensory neuron, in these animals caused aggregation, bordering and locomotion speed to return to nearly the levels seen in social strains, suggesting that ASK can promote social feeding.

The next obvious question was how the ASK–RMG spoke of the circuit might act to regulate foraging style. ASK is required to attract males to hermaphrodites⁸, and Macosko and colleagues show that it signals the presence of pheromones by a reduction in its activity. These clues prompted the authors to test whether social and solitary strains might differ in their behavioural responses to pheromones. Whereas pheromones repelled solitary animals (as previously shown^{8,9}), they attracted the *npr-1*-defective social strains. This striking observation has two key implications. First, it points to pheromones as the missing, long-range sensory cue that attracts worms into feeding groups. Second, it brings *C. elegans* aggregation closer to the orbit of behaviours that are more than just metaphorically social.

Macosko *et al.* also find that a pulse of pheromones reduces ASK activity to a much greater extent in the *npr-1*-defective social strain than in the solitary strain. This observation provides a clue as to how the RMG–ASK spoke of the circuit might function. In one model, RMG resides in either a low (solitary feeding) or a high (social feeding) activity state, depending on the activity of NPR-1 and, possibly, on the concentration of neuropeptides (Fig. 1b). The state of RMG activity is communicated electrically to ASK through gap junctions, with the result that membrane potential in ASK is set respectively low or high. When ASK is in its low state, there is little activity to inhibit. But when it is in its high state, a pheromone pulse can more strongly inhibit it. Thus, RMG switches the dynamic range of ASK's pheromone response.

In support of this model, the researchers⁴ show that pheromone-induced synaptic responses in the interneuron AIA — the main postsynaptic target of ASK — are small in the solitary strain and large in the social strain.

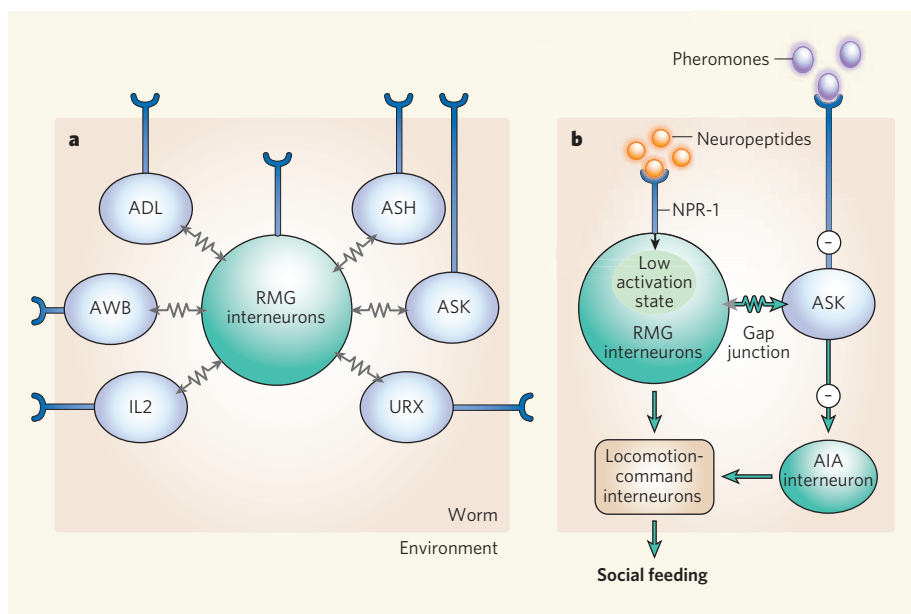


Figure 1 | The RMG hub-and-spoke circuit. **a**, RMG interneurons are connected to six types of sensory neuron (ASH, ASK, URX, IL2, AWB and ADL) covering a wide range of sensory modalities. The circuit might operate in two directions simultaneously. In the centripetal direction, RMG could integrate signals across several modalities to assess the quality of the local environment. In the centrifugal direction, RMG could set the dynamic range of the sensory neurons' response. **b**, For instance, Macosko and colleagues' data⁴ indicate that RMG interneurons might send electrical signals through gap junctions to ASK, tuning the extent of this sensory neuron's response to pheromones. When NPR-1 is absent or its activity is low, RMG activation is high, leading to ASK's strong response to pheromones and so its relatively strong signals to locomotion-command interneurons through AIA interneurons, which might promote social feeding. Negative signs indicate inhibition; all other connections are positive.

Thus, from the point of view of the neurons in the social strain that get their news about the proximity of other nematodes from ASK, it would seem as though the animal has encountered a bigger change in pheromone concentration, presumably leading to a stronger orientation response. It would be interesting to determine whether AIA is required for aggregation responses.

An attractive feature of the hub-and-spoke motif is that, if its gap junctions pass ionic currents bidirectionally (as many do), then the circuit ought to operate simultaneously in opposing directions. In the centrifugal direction, RMG should broadcast to sensory neurons of various modalities its assessment of the animal's internal state, encoded by neuropeptides, thereby redirecting the attentions of the animal according to its needs. In the centripetal direction, by simple laws of current flow in circuits, RMG's membrane voltage should be a weighted average of the state of the sensory neurons, providing a balanced assessment of the local environment. We do not yet know whether the hub-and-spoke circuit as a whole functions as proposed. The evidence that one of the seven spokes (RMG–ASK) functions centrifugally now provides the impetus to test other spokes as well.

As neuroscientists contemplate large-scale anatomical reconstructions of other nervous systems¹⁰, it can be argued that costly wiring diagrams are not worth the expense because

of unavoidable ambiguities in the final result. Macosko and colleagues' work⁴ provides a bracing example to the contrary, for without the guidance of the *C. elegans* wiring diagram⁶, one might have lost sight of this rich vein of inquiry.

Shawn R. Lockery is at the Institute of Neuroscience, 1254 University of Oregon, Eugene, Oregon 97403, USA.
e-mail: shawn@chinook.uoregon.edu

1. de Bono, M. & Bargmann, C. I. *Cell* **94**, 679–689 (1998).
2. Gray, J. M. *et al. Nature* **430**, 317–322 (2004).
3. Persson, A. *et al. Nature* doi:10.1038/nature07820 (2009).
4. Macosko, E. Z. *et al. Nature* **458**, 1171–1175 (2009).
5. Coates, J. C. & de Bono, M. *Nature* **419**, 925–929 (2002).
6. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
7. Schackwitz, W. S., Inoue, T. & Thomas, J. H. *Neuron* **17**, 719–728 (1996).
8. Srinivasan, J. *et al. Nature* **454**, 1115–1118 (2008).
9. White, J. Q. *et al. Curr. Biol.* **17**, 1847–1857 (2007).
10. Helmstaedter, M., de Kock, C. P. J., Feldmeyer, D., Bruno, R. M. & Sakmann, B. *Brain Res. Rev.* **55**, 193–203 (2007).

Corrections

● In the News & Views article “Quantum chemistry: The little molecule that could” by Chris H. Greene (*Nature* **458**, 975–976; 2009), the equivalent for 1 millikelvin should have been given as 100 billionths of an electronvolt, not 4 billionths.

● In one of the obituary articles about John Maddox, “Maddox by his successor” by Philip Campbell (*Nature* **458**, 985–986; 2009), Nietzsche was misspelled.

Cytoplasmic functions of the tumour suppressor p53

Douglas R. Green^{1,2} & Guido Kroemer^{3,4}

The principal tumour-suppressor protein, p53, accumulates in cells in response to DNA damage, oncogene activation and other stresses. It acts as a nuclear transcription factor that transactivates genes involved in apoptosis, cell cycle regulation and numerous other processes. An emerging area of research unravels additional activities of p53 in the cytoplasm, where it triggers apoptosis and inhibits autophagy. These previously unknown functions contribute to the mission of p53 as a tumour suppressor.

Transcriptional and non-transcriptional effects of p53

Approximately half of human cancers have inactivating mutations of p53 (known as TP53 in human), and most of the remaining malignancies deactivate the p53 pathway by increasing its inhibitors, reducing its activators or inactivating its downstream targets. p53 is best characterized as a transcription factor that binds to specific DNA sequences and transactivates a number of genes with a variety of functions including cell cycle arrest, apoptosis, causing changes in metabolism and others¹ (Fig. 1). In addition to this nuclear activity, p53 also possesses biological activities that are cytosolic and transcription-independent. Several years ago it was noted² that overexpression of a mutant p53, lacking most of the DNA-binding domain (DBD) and completely deficient in transactivation function, could nonetheless trigger apoptosis. Indeed, overexpression of a variety of transactivation-incompetent p53 mutants can efficiently induce apoptosis in human cells³. Consistent with this, it was found⁴ that apoptosis induced by stabilization of an ectopically expressed temperature-sensitive mutant of p53 could proceed in the absence of RNA and protein synthesis. Similarly, activation of p53 was found to trigger apoptosis even in the absence of a nucleus⁵. p53-reactivating

drugs that interact with oncogenic, mutant p53 protein causing it to adopt a wild-type conformation can induce apoptosis under conditions of complete transcriptional or translational blockade^{5,6}. Most recently, mice were generated in which the endogenous p53 (also known as *Trp53* in mouse) gene was replaced by a chimaeric p53 protein that is capable of transactivation yet lacks several domains that are required for other p53 functions⁷. In fibroblasts from such mice, this chimaeric p53 protein was transcriptionally active and able to induce cellular senescence, but was unable to trigger apoptosis⁷. Together, these observations support the idea that a cytoplasmic pool of p53 can induce apoptosis through a transactivation-independent mechanism.

p53 is at the hub of numerous signalling pathways triggered by a range of cellular stresses including DNA damage by exogenous mutagens, oncogene activation, telomere erosion and hypoxia, all of which influence the abundance, subcellular localization, post-translational modification and/or interaction of p53 with cofactors. As a result of these context-dependent, damage-elicited alterations, p53 can facilitate the transient adaptation of cells to stressful conditions, for example, by increasing DNA repair upon a transient cell cycle arrest or by enhancing the expression of enzymes that detoxify reactive oxygen species. Alternatively, p53 suppresses oncogenic potential by mediating an irreversible arrest of the cell cycle or by triggering apoptotic cell death⁸. In addition to its role as a tumour suppressor, p53 has a major role in ageing⁹, as well as in the unwarranted loss of post-mitotic cells such as heart muscle cells in infarction and neurons in stroke¹⁰. It is likely that the biological effects of p53 represent the combined activities of the nuclear and cytoplasmic protein.

Effects of cytoplasmic p53 on mitochondria

The mitochondrial membrane constitutes the battleground on which pro- and anti-apoptotic factors induce or prevent a potentially lethal permeabilization step¹¹. Under a variety of cell-death-inducing conditions, p53 rapidly moves to the mitochondria. For instance, whole-body irradiation of mice causes a fraction of cellular p53 to associate with the outer mitochondrial membrane¹². Similarly, ischaemic damage of the rat brain triggers the translocation of p53 (also known as Tp53 in rat) to the mitochondria of neurons that are particularly vulnerable to hypoxia within the CA1 area of the hippocampus¹⁰. Once at the mitochondrion, p53 induces mitochondrial outer membrane permeabilization (MOMP), thereby triggering the release of pro-apoptotic factors from the mitochondrial intermembrane space.

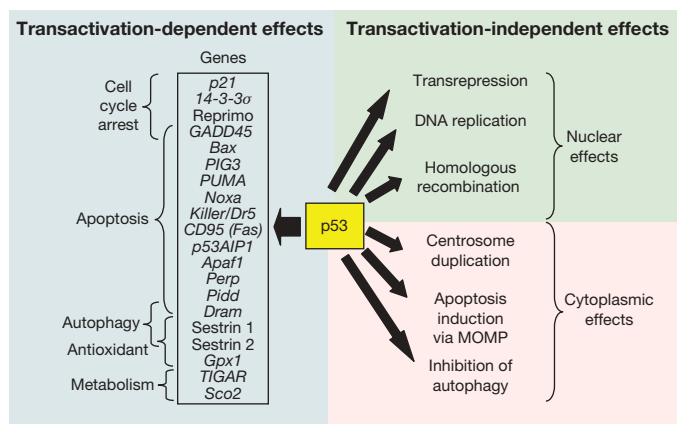


Figure 1 | Classification of p53 activities. On the left side, some genes that are transactivated by p53 are exemplified, together with a few of the functional consequences of p53 activation. On the right side, transactivation-independent effects of p53 are listed. These can be divided into nuclear and extra-nuclear (cytoplasmic) p53 activities.

¹Department of Immunology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ²Institut Gustave Roussy, ³Université Paris Sud, Paris 11, ⁴INSERM, U848, F-94805 Villejuif, France.

Indeed, mitochondrion-targeted p53 can be as efficient as wild-type p53 in inhibiting tumour growth¹².

Multiple mechanisms have been invoked to explain how p53 triggers MOMP¹². MOMP is usually inhibited by anti-apoptotic multidomain proteins of the Bcl2 family (such as Bcl2, Bcl-X_L (also known as Bcl2l1) and Mcl1), and is conditional on pro-apoptotic multidomain proteins from the same family (in particular Bax and Bak (Bak1)) that can homo-oligomerize within the outer mitochondrial membrane to form MOMP-mediated supramolecular structures. Depending on their particular affinities for multidomain Bcl2 family proteins, a set of distinct pro-apoptotic 'BH3-only' proteins can directly interact with Bax or Bak to trigger their homo-oligomerization and hence MOMP (these BH3-only proteins are referred to as 'direct activators') and/or neutralize one or more anti-apoptotic multidomain proteins (referred to as 'sensitizers' or 'de-repressors'). In contrast to direct activators (in particular the proteins Bim (Bcl2l11) and Bid), sensitizers cannot trigger MOMP on their own (because they simply antagonize its inhibition) and require the input of additional stimuli for apoptosis induction¹³.

p53 has been suggested to act like a BH3-only protein, either as a direct activator of Bax and/or Bak or as a de-repressor. Under pro-apoptotic conditions, p53 can be co-immunoprecipitated with Bcl2, Bcl-X_L and Bak^{12,14,15}. In a defined system involving only recombinant proteins and synthetic membranes, p53 can trigger Bax to permeabilize liposomes through a 'hit and run' mechanism—that is, through transient molecular associations¹⁶. NMR structures of p53 in a complex with Bcl2 or Bcl-X_L indicate that the DBD of p53 is involved in docking with Bcl2 family proteins^{12,17}, further supported by biochemical studies on p53 and Bak¹⁸. Hence, oncogenic p53 mutations affecting the DBD can operate as 'dual hits' and simultaneously abrogate transactivation and direct MOMP induction by p53. In addition, other studies show that the proline-rich region neighbouring the DBD is important for Bax activation¹⁶ and can also associate with anti-apoptotic Bcl2 proteins¹⁸. Further, the p53-binding interfaces within Bcl-X_L and Bak may be distinct, as determined by biochemical approaches¹⁷. Because p53 has a higher affinity for Bcl-X_L than for Bak¹⁹, it may first engage in molecular interactions with anti-apoptotic Bcl2 proteins (as a sensitizer) and then with Bak and Bax (as a direct activator). Another nuclear transcription factor, Nur77, has previously been shown to bind to Bcl2 in a way that radically changes its conformation, apparently converting Bcl2 into a pro-apoptotic protein²⁰. However, the structural changes that p53 imposes on Bcl2 or Bcl-X_L are minor^{12,19}, indicating that p53 does not function in this manner.

The pro-apoptotic effects of cytoplasmic p53 are not dependent on transcription, in principle. However, the control of transcription by nuclear p53 decisively contributes to the function of cytoplasmic p53. As discussed in more detail below, the p53 target Mdm2 is essential for post-translational regulation of p53, without which the system would not be responsive to cellular stress. Another p53 target, PUMA (Bbc3), controls the sequestration of cytoplasmic p53 by the anti-apoptotic Bcl-X_L protein, releasing p53 to activate Bax²¹. Therefore, without transcription, regulated by nuclear p53, endogenous cytoplasmic p53 may not function (Fig. 2).

One p53-targeted drug, pifithrin- μ , inhibits the pro-apoptotic effects of cytosolic p53, but has no apparent effect on p53-dependent transactivation²². This drug blocks the interaction of p53 with Bcl-X_L²², and probably also the interaction of p53 with pro-apoptotic Bcl2 family members (Bax and Bak), thereby accounting for its anti-apoptotic effects. Pifithrin- μ can rescue mice from otherwise lethal irradiation²², indicating that selective inhibition of the cytoplasmic p53 pathway is sufficient for radioprotection *in vivo*. Conversely, p53-activating drugs such as CP-31398 induce p53 translocation to mitochondria, as well as p53-dependent MOMP. This p53 translocation can be inhibited by cyclosporin A (CsA)⁶, pointing to a possible molecular crosstalk between the CsA-inhibitable mitochondrial permeability transition pore and the p53 system. Reportedly,

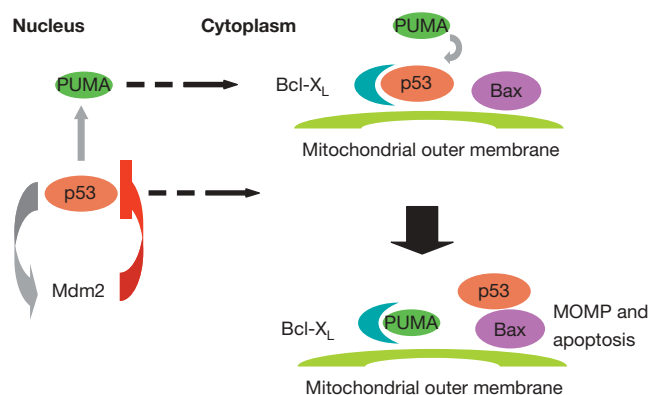


Figure 2 | Interplay of the nuclear and cytoplasmic functions of p53 in apoptosis. Nuclear p53 induces the expression of Mdm2, which acts to inhibit the protein through binding and ubiquitinylation. Cellular stress signals interrupt this inhibition, allowing p53 to accumulate both in the nucleus and in the cytoplasm. In the latter, p53 is sequestered by anti-apoptotic Bcl2 proteins such as Bcl-X_L. Another target of nuclear p53, PUMA, functions to disrupt the Bcl-X_L-p53 interaction. The released p53 can now trigger MOMP and apoptosis through interaction with, for example, Bax.

recombinant p53 may induce a more complete MOMP (with release of the pro-apoptotic factor AIF (Aifm1)) than recombinant Bid added to purified mitochondria *in vitro*²³, pointing to possible differences in mitochondrial permeabilization by p53 and BH3-only proteins.

Inhibition of autophagy by cytoplasmic p53

Macroautophagy (referred to as autophagy) is the sequestration and subsequent digestion of parts of the cytoplasm, allowing for the adaptation of cells to stressful conditions, as well as the removal of damaged, potentially harmful cytoplasmic organelles. Enhanced autophagy, which frequently accompanies cell death, likewise constitutes a failed attempt to adapt to stress and to survive, rather than a lethal catabolic process²⁴. Because autophagy has an essential role in the maintenance of genomic stability²⁵, inhibition of autophagy is oncogenic. Accordingly, loss of only one allele of either of the two haploinsufficient autophagy genes beclin 1 (*Becn1*) or *Uvrag* is sufficient to promote carcinogenesis, and multiple oncogenes including Bcl2, Akt (Akt1) and PI(3)K inhibit autophagy. Similarly, the inactivation of tumour-suppressor proteins such as Pten, Tsc1, Tsc2 and Lkb1 (also known as Stk11) results in autophagy inhibition²⁴.

Although p53 can transactivate genes that induce autophagy (such as DRAM and sestrins 1 and 2)^{26,27}, normal levels of p53 mediate a tonic inhibition of autophagy (Fig. 1). In fact, the deletion, depletion or pharmacological inhibition of p53 with pifithrin- α induces autophagy in mouse, human and nematode cells²⁸. Suppression of autophagy is mediated by cytoplasmic, not nuclear, p53, and physiological inducers of autophagy (such as nutrient depletion) must destroy the pool of cytoplasmic p53 to induce autophagy²⁸. Thus, inhibition of the ubiquitin E3 ligase Mdm2, which targets p53 for destruction, can suppress the induction of autophagy by starvation, rapamycin, lithium or damage of the endoplasmic reticulum²⁸. Cytoplasmic (but not nuclear) p53 inhibits the AMP-dependent kinase, a positive regulator of autophagy, and activates mammalian target of rapamycin (mTOR, also known as Frap1), a negative regulator of autophagy²⁸. How these effects are achieved, however, remains an open conundrum. Nevertheless, it is not unlikely that the transactivation-dependent metabolic effects of p53 and its cytosolic, transcription-independent inhibition of autophagy cooperate to ensure a coordinated action of p53 in cellular adaptation, such as in reprogramming metabolism towards oxidative phosphorylation.

It is tempting to speculate that the dual action of cytoplasmic p53— inhibition of autophagy and induction of MOMP—may constitute a coordinated response for cell death induction (Fig. 3). Autophagy

accounts for the removal of damaged and permeabilized mitochondria and counteracts the lethal effect of MOMP²⁹. Therefore, autophagy inhibition by p53 may further facilitate cell death execution by MOMP. Nonetheless, it is not clear through which mechanisms cytoplasmic p53 switches from its baseline function (autophagy inhibition) to its killer activity (translocation to mitochondria and MOMP induction).

At first glance it appears paradoxical that normal levels of cytoplasmic p53 inhibit autophagy, although inhibition of autophagy is often associated with oncogenesis. This paradox is resolved by the observation that mutant p53 protein, which has lost its transactivation function and which accumulates in the cytoplasm of tumour cells, efficiently inhibits autophagy³⁰. The structural features of p53 required for its cytoplasmic pro-apoptotic and anti-autophagic functions are clearly distinct. Indeed, deletion of the DBD does not affect autophagy inhibition by p53. Moreover, point mutations that affect the nuclear functions of p53, as well as its interaction with Bcl2 family proteins, do not abolish its capacity to inhibit autophagy³⁰. This may contribute to the strong oncogenic action of certain p53 mutants that is difficult to explain by the mere abolition of their tumour-suppressive capabilities.

Regulating the regulator

The nuclear versus cytoplasmic effects of p53 are determined by multiple post-translational modifications that affect its interaction with other proteins, its shuttling between the cytoplasm and the nucleus and its biological activities. Poly(ADP)ribosylation of p53 leads to its nuclear accumulation⁸. In contrast, monoubiquitylation by Mdm2 stimulates the nuclear export of p53, which on arrival at mitochondria is deubiquitylated by mitochondrial HAUSP, thus generating the apoptotically active non-ubiquitylated p53 (ref. 31). Other post-translational modifications of p53 (such as phosphorylation of carboxy-terminal serines) can stimulate nuclear export and/or mitochondrial association. Moreover, the transcription factor Foxo3a (Foxo3) promotes p53 cytoplasmic accumulation by increasing its nuclear export, hence stimulating direct, p53-mediated MOMP induction³². This indicates that the entire context of post-transcriptional p53 modifications and protein interactions can affect the precise subcellular localization and function of p53.

A previously unknown class of p53 activators, the tenovins, activate the tumour-suppressive function of p53 by inhibiting the sirtuins Sirt1 and Sirt2 (ref. 33)—p53 deacetylases—in tumour cells. Similarly, a tumour suppressor, Dbc1 (deleted in breast cancer 1), acts as an endogenous inhibitor of Sirt1 and as a positive regulator of p53 (ref. 34). However, according to one report³⁵, nuclear p53 is

acetylated whereas cytoplasmic p53 is deacetylated. Knockout of Sirt1 facilitates the reactive-oxygen-species-induced nuclear translocation of p53 and simultaneously inhibits direct MOMP induction by cytoplasmic p53, at least in mouse embryonic stem cells³⁵. Future investigations must resolve this apparent contradiction to understand which post-transcriptional modifications of p53 determine its pro-MOMP and anti-autophagy activities.

Future directions and perspectives

Cancer cells are characterized by failing cell cycle checkpoints, reduced propensity to apoptosis and suppressed autophagy. 'Hotspot' mutations of p53 within the DBD usually abolish the transactivation function of p53 (and often create dominant-negative inhibitors of wild-type p53, with which they form heterotetramers), thereby preventing the expression of cell-cycle-arresting, pro-apoptotic and autophagy-inducing genes. Such p53 mutations within the DBD also affect the cytoplasmic functions of p53, reducing its capacity to induce MOMP (through failure to interact with Bcl2 family proteins)^{12,15,17}, yet leaving intact its inhibitory effect on autophagy³⁰. Whether and how such mutations perturb the cytoplasmic regulation of cell cycle checkpoints remains elusive. However, it appears that frequent p53 mutations can contribute to oncogenesis through the concerted subversion of both the nuclear and cytoplasmic programs of tumour suppression (Fig. 3). Similarly, the inactivation of the oncosuppressor Arf (Cdkn2a; which, through the activation of Mdm2, leads to p53 depletion) combines transcriptional effects (due to the absence of p53) with cytoplasmic ones. Indeed, distinct splice variants of Arf can induce MOMP (though through a mechanism distinct from that of p53, involving an interaction with a specific mitochondrial receptor, p32 (C1qbp), or an interaction with Bcl-X_L) and stimulate autophagy^{36–38}. Therefore, the net result of Arf inactivation may also be a combined subversion of apoptosis and autophagy, both at the nuclear and at the extra-nuclear levels. ARC (Nol3), an apoptosis-inhibitory protein that is overexpressed in numerous cancers, is present in the nucleus, in which it inhibits p53 tetramerization and stimulates its export, as well as in the cytoplasm, where it neutralizes Bax to inhibit cell death³⁹. It remains an open question through which mechanisms many other oncogenic perturbations, such as constitutive activation of the insulin receptor pathway, may affect (p53-dependent?) tumour suppression in the cytoplasm. Understanding the extra-nuclear activities of p53 will likewise furnish new opportunities to pharmacologically modulate the p53 system.

p53 has a prominent—and controversial—role in the regulation of ageing and longevity⁹. In the nematode *Caenorhabditis elegans*, knock-out of the p53 orthologue *cep-1* fails to cause oncogenesis, yet significantly increases both median and maximum lifespan. This gain of longevity is lost when autophagy is inhibited⁴⁰. We anticipate that the investigation of whether and how p53 can participate in a longevity-increasing pathway that links apoptosis, caloric restriction, activation of sirtuins and regulation of autophagy will yield crucial insights into the intricate relationship between tumour suppression and ageing that dictates our inexorable, yet variable, fate.

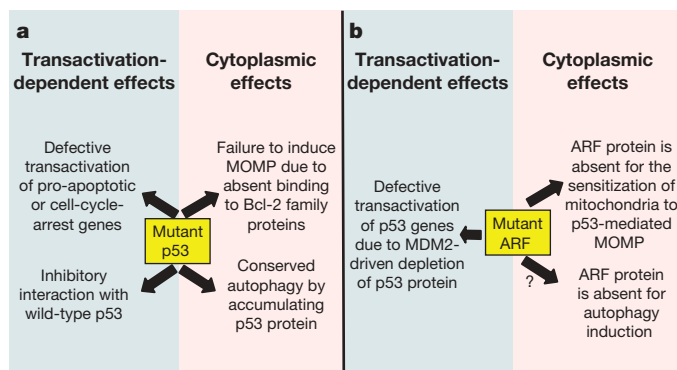


Figure 3 | Concerted oncogenic actions of mutant p53 or inactive ARF in the nucleus and cytoplasm of cancer cells. Hotspot mutations of p53 affecting the DBD can abolish the transactivation of p53 genes as well as the mitochondrion-permeabilizing action of p53, yet leave intact autophagy inhibition by p53 (a). Similarly, oncogenic mutations that affect the C terminus of ARF can lead to the depletion of p53 protein, as well as to the abolition of mitochondrion-permeabilizing and autophagy-inducing activities mediated by the ARF protein (b).

1. Riley, T., Sontag, E., Chen, P. & Levine, A. Transcriptional control of human p53-regulated genes. *Nature Rev. Mol. Cell Biol.* **9**, 402–412 (2008).
2. Haupt, Y., Rowan, S., Shaulian, E., Vousden, K. H. & Oren, M. Induction of apoptosis in HeLa cells by trans-activation-deficient p53. *Genes Dev.* **9**, 2170–2183 (1995).
3. Kakudo, Y., Shibata, H., Otsuka, K., Kato, S. & Ishioka, C. Lack of correlation between p53-dependent transcriptional activity and the ability to induce apoptosis among 179 mutant p53s. *Cancer Res.* **65**, 2108–2114 (2005).
4. Caelles, C., Helmborg, A. & Karin, M. p53-dependent apoptosis in the absence of transcriptional activation of p53-target genes. *Nature* **370**, 220–223 (1994).
5. Chipuk, J. E., Maurer, U., Green, D. R. & Schuler, M. Pharmacologic activation of p53 elicits Bax-dependent apoptosis in the absence of transcription. *Cancer Cell* **4**, 371–381 (2003).
6. Tang, X. *et al.* CP-31398 restores mutant p53 tumor suppressor function and inhibits UVB-induced skin carcinogenesis in mice. *J. Clin. Invest.* **117**, 3753–3764 (2007).

7. Johnson, T. M., Meade, K., Pathak, N., Marques, M. R. & Attardi, L. D. Knockin mice expressing a chimeric p53 protein reveal mechanistic differences in how p53 triggers apoptosis and senescence. *Proc. Natl Acad. Sci. USA* **105**, 1215–1220 (2008).
8. Murray-Zmijewski, F., Slee, E. A. & Lu, X. A complex barcode underlies the heterogeneous response of p53 to stress. *Nature Rev. Mol. Cell Biol.* **9**, 702–712 (2008).
9. Finkel, T., Serrano, M. & Blasco, M. A. The common biology of cancer and ageing. *Nature* **448**, 767–774 (2007).
10. Endo, H., Saito, A. & Chan, P. H. Mitochondrial translocation of p53 underlies the selective death of hippocampal CA1 neurons after global cerebral ischaemia. *Biochem. Soc. Trans.* **34**, 1283–1286 (2006).
11. Kroemer, G., Galluzzi, L. & Brenner, C. Mitochondrial membrane permeabilization in cell death. *Physiol. Rev.* **87**, 99–163 (2007).
12. Moll, U. M., Marchenko, N. & Zhang, X. K. p53 and Nur77/TR3 — transcription factors that directly target mitochondria for cell death induction. *Oncogene* **25**, 4725–4743 (2006).
13. Letai, A. G. Diagnosing and exploiting cancer's addiction to blocks in apoptosis. *Nature Rev. Cancer* **8**, 121–132 (2008).
14. Leu, J. I., Dumont, P., Hafey, M., Murphy, M. E. & George, D. L. Mitochondrial p53 activates Bak and causes disruption of a Bak–Mcl1 complex. *Nature Cell Biol.* **6**, 443–450 (2004).
15. Mihara, M. *et al.* p53 has a direct apoptogenic role at the mitochondria. *Mol. Cell* **11**, 577–590 (2003).
16. Chipuk, J. E. *et al.* Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. *Science* **303**, 1010–1014 (2004).
This work characterizes protein–protein interactions between p53 and pro-apoptotic Bcl2 proteins as the mechanism through which p53 can trigger the formation of supramolecular, protein-permeable conduits in the outer mitochondrial membrane.
17. Pietsch, E. C. *et al.* Oligomerization of BAK by p53 utilizes conserved residues of the p53 DNA binding domain. *J. Biol. Chem.* **283**, 21294–21304 (2008).
This study uses biochemical approaches to map the interaction of p53 with Bak to a region corresponding to the analogous Bim–Bax interaction site.
18. Xu, H., Tai, J., Ye, H., Kang, C. B. & Yoon, H. S. The N-terminal domain of tumor suppressor p53 is involved in the molecular interaction with the anti-apoptotic protein Bcl-Xl. *Biochem. Biophys. Res. Commun.* **341**, 938–944 (2006).
19. Sot, B., Freund, S. M. & Fersht, A. R. Comparative biophysical characterization of p53 with the pro-apoptotic BAK and the anti-apoptotic BCL-xL. *J. Biol. Chem.* **282**, 29193–29200 (2007).
20. Lin, B. *et al.* Conversion of Bcl-2 from protector to killer by interaction with nuclear orphan receptor Nur77/TR3. *Cell* **116**, 527–540 (2004).
21. Chipuk, J. E., Bouchier-Hayes, L., Kuwana, T., Newmeyer, D. D. & Green, D. R. PUMA couples the nuclear and cytoplasmic proapoptotic function of p53. *Science* **309**, 1732–1735 (2005).
22. Strom, E. *et al.* Small-molecule inhibitor of p53 binding to mitochondria protects mice from gamma radiation. *Nature Chem. Biol.* **2**, 474–479 (2006).
23. Wolff, S., Erster, S., Palacios, G. & Moll, U. M. p53's mitochondrial translocation and MOMP action is independent of Puma and Bax and severely disrupts mitochondrial membrane integrity. *Cell Res.* **18**, 733–744 (2008).
24. Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).
25. Mathew, R., Karantza-Wadsworth, V. & White, E. Role of autophagy in cancer. *Nature Rev. Cancer* **7**, 961–967 (2007).
26. Crighton, D. *et al.* DRAM, a p53-induced modulator of autophagy, is critical for apoptosis. *Cell* **126**, 121–134 (2006).
27. Budanov, A. V. & Karin, M. p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. *Cell* **134**, 451–460 (2008).
28. Tasdemir, E. *et al.* Regulation of autophagy by cytoplasmic p53. *Nature Cell Biol.* **10**, 676–687 (2008).
This article reports the discovery that cytoplasmic p53 can inhibit autophagy in several species (humans, mice and nematodes), thus revealing a novel crosstalk between autophagy and apoptosis.
29. Colell, A. *et al.* GAPDH and autophagy preserve survival after apoptotic cytochrome c release in the absence of caspase activation. *Cell* **129**, 983–997 (2007).
30. Morselli, E. *et al.* Mutant p53 protein localized in the cytoplasm inhibits autophagy. *Cell Cycle* **7**, 3056–3061 (2008).
31. Marchenko, N. D., Wolff, S., Erster, S., Becker, K. & Moll, U. M. Monoubiquitylation promotes mitochondrial p53 translocation. *EMBO J.* **26**, 923–934 (2007).
32. You, H., Yamamoto, K. & Mak, T. W. Regulation of transactivation-independent proapoptotic activity of p53 by FOXO3a. *Proc. Natl Acad. Sci. USA* **103**, 9051–9056 (2006).
This study establishes a role for the transcription factor Foxo3a in accumulation and function of cytosolic p53.
33. Lain, S. *et al.* Discovery, *in vivo* activity, and mechanism of action of a small-molecule p53 activator. *Cancer Cell* **13**, 454–463 (2008).
34. Zhao, W. *et al.* Negative regulation of the deacetylase SIRT1 by DBC1. *Nature* **451**, 587–590 (2008).
35. Han, M. K. *et al.* SIRT1 regulates apoptosis and Nanog expression in mouse embryonic stem cells by controlling p53 subcellular localization. *Cell Stem Cell* **2**, 241–251 (2008).
36. Reef, S. *et al.* A short mitochondrial form of p19ARF induces autophagy and caspase-independent cell death. *Mol. Cell* **22**, 463–475 (2006).
37. Itahana, K. & Zhang, Y. Mitochondrial p32 is a critical mediator of ARF-induced apoptosis. *Cancer Cell* **13**, 542–553 (2008).
38. Pimkina, J., Humbey, O., Zilfou, J. T., Jarnik, M. & Murphy, M. ARF induces autophagy by virtue of interaction with Bcl-xl. *J. Biol. Chem.* **284**, 2803–2810 (2008).
39. Foo, R. S. *et al.* Regulation of p53 tetramerization and nuclear export by ARC. *Proc. Natl Acad. Sci. USA* **104**, 20826–20831 (2007).
40. Tavernarakis, N., Pasparaki, A., Tasdemir, E., Maiuri, M. C. & Kroemer, G. The effects of p53 on whole organism longevity are mediated by autophagy. *Autophagy* **4**, 870–873 (2008).

Acknowledgments The authors' own work is supported by NIH and the American Lebanese and Syrian Associated Charities (to D.R.G.) and by Ligue contre le Cancer, INCa, Cancéropole, ANR, ANRS and the Active p53 and Apo-Sys EU networks (to G.K.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R.G. (douglas.green@stjude.org) or G.K. (kroemer@igr.fr).

Autophagy regulates lipid metabolism

Rajat Singh^{1,2*}, Susmita Kaushik^{1,2,3,4*}, Yongjun Wang^{1,2}, Youqing Xiang^{1,2}, Inna Novak^{2,5}, Masaaki Komatsu⁶, Keiji Tanaka⁶, Ana Maria Cuervo^{1,2,3,4} & Mark J. Czaja^{1,2}

The intracellular storage and utilization of lipids are critical to maintain cellular energy homeostasis. During nutrient deprivation, cellular lipids stored as triglycerides in lipid droplets are hydrolysed into fatty acids for energy. A second cellular response to starvation is the induction of autophagy, which delivers intracellular proteins and organelles sequestered in double-membrane vesicles (autophagosomes) to lysosomes for degradation and use as an energy source. Lipolysis and autophagy share similarities in regulation and function but are not known to be interrelated. Here we show a previously unknown function for autophagy in regulating intracellular lipid stores (macrolipophagy). Lipid droplets and autophagic components associated during nutrient deprivation, and inhibition of autophagy in cultured hepatocytes and mouse liver increased triglyceride storage in lipid droplets. This study identifies a critical function for autophagy in lipid metabolism that could have important implications for human diseases with lipid over-accumulation such as those that comprise the metabolic syndrome.

Free fatty acids (FFAs) are taken up by hepatocytes and converted into triglycerides (TGs) for storage with cholesterol in lipid droplets (LDs)¹. LD-sequestered TGs continually undergo hydrolysis, generating FFAs that are predominantly re-esterified back into TGs for storage^{1,2}. Nutrient deprivation upregulates TG hydrolysis to supply FFAs for oxidation to meet cellular energy demands³. An alternative energy source in times of nutrient scarcity is provided by the breakdown of cellular components by autophagy^{4,5}. Both macroautophagy (the type of autophagy quantitatively more important and subsequently referred to as autophagy) and lipolysis are regulated hormonally by insulin and glucagon⁶ and are increased during starvation. Except for

the processing of endocytosed lipoproteins, no direct involvement of the lysosomal degradation pathway in lipid metabolism has been established. The regulatory and functional similarities between autophagy and lipolysis, along with the capability of lysosomes to degrade lipids, indicated that autophagy may contribute to LD and TG breakdown (Supplementary Fig. 1).

Inhibition of autophagy increases lipid storage

Pharmacological inhibition of autophagy with 3-methyladenine (3MA)⁷ significantly increased hepatocyte TG content in the absence or presence of exogenous lipid supplementation with oleate (Fig. 1a).

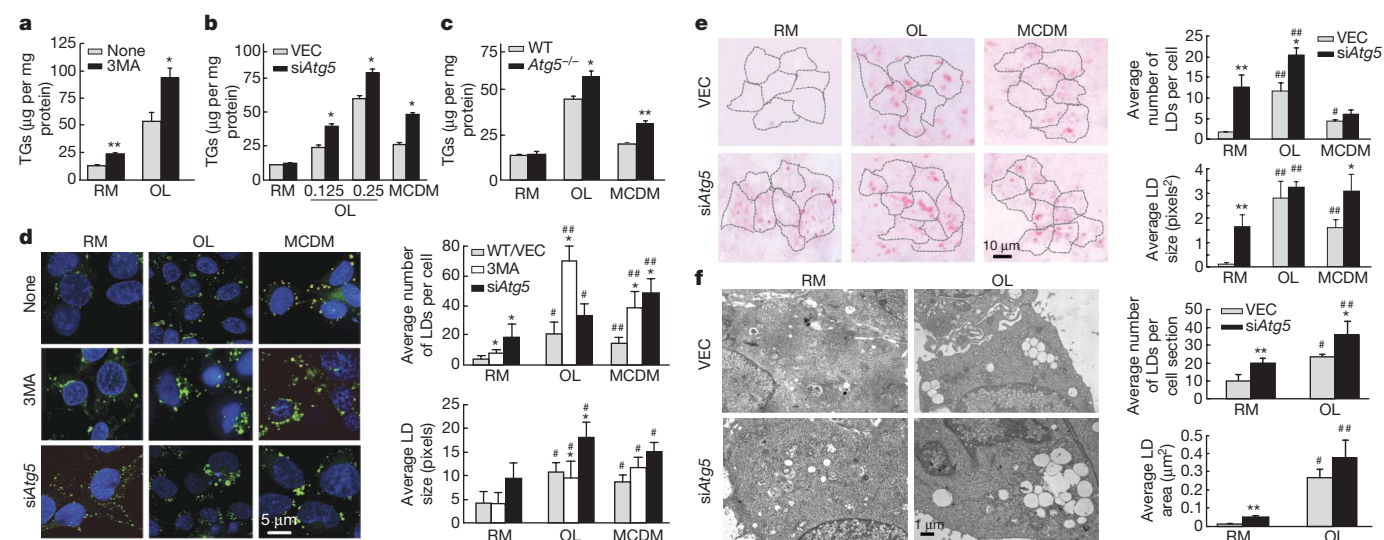


Figure 1 | Inhibition of autophagy leads to increased TG accumulation.

a, TG levels in hepatocytes untreated (None) or treated with 3-methyladenine (3MA) and cultured in regular medium (RM) or oleate (OL) (* $P < 0.02$, ** $P < 0.002$, $n = 3$). **b**, TG levels in vector-infected (VEC) and siAtg5 cells in RM, OL or in MCDM (* $P < 0.001$, $n = 5$). OL values are in mM. **c**, TG levels in wild-type (WT) or Atg5 knockout mice embryonic

fibroblasts (Atg5^{-/-}) (* $P < 0.01$ or ** $P < 0.0001$, $n = 5$). **d–f**, Cells from **a** and **b** stained with BODIPY 493/503 (**d**), oil red O (**e**) or visualized by electron microscopy (**f**). Right: quantifications of LD number and size (* $P < 0.01$, ** $P < 0.001$ with untreated wild-type cells; # $P < 0.01$, ## $P < 0.001$ with cells in RM; $n = 4–6$). Nuclei are highlighted with 4,6-diamidino-2-phenylindole (DAPI; **d**). Error bars, s.e.m.

¹Department of Medicine, ²The Marion Bessin Liver Research Center, ³Department of Developmental and Molecular Biology, ⁴Institute for Aging Studies, ⁵Department of Pediatrics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. ⁶Laboratory of Frontier Science, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113-8613, Japan.

*These authors contributed equally to this work.

A knockdown of the autophagy gene *Atg5* in hepatocytes (siAtg5 cells; Supplementary Fig. 2a) also increased TG levels with oleate or a second endogenous stimulus for TG formation—culture in methionine- and choline-deficient medium (MCDM)^{8,9} (Fig. 1b and Supplementary Fig. 2b with a second short hairpin RNA, shRNA). TG levels were also increased in embryonic fibroblasts from *Atg5* knockout mice (Fig. 1c). Although cellular cholesterol content was unaffected in siAtg5 cells by oleate alone or culture in MCDM, oleate and cholesterol co-treatment led to significantly greater cholesterol content (Supplementary Fig. 2c, d).

Consistent with the increased TG levels, lipid staining with BODIPY 493/503 or oil red O revealed increased LD number and size in hepatocytes with oleate or MCDM that were further increased by addition of 3MA or *Atg5* knockdown (Fig. 1d, e). These increases were greater than for TG levels because the lipid stains detect all neutral lipids in LDs (TGs and cholesterol). Similar results were obtained with palmitate (Supplementary Fig. 3). An increase in the number and size of LDs in oleate-treated siAtg5 cells was also confirmed by electron microscopy (Fig. 1f). The absence of co-localization of BODIPY 493/503 and an endoplasmic reticulum marker (Supplementary Fig. 4), and the co-localization of lipid with the LD-associated protein TIP47 (also known as M6PRBP1; ref. 1, Supplementary Fig. 5), demonstrated that lipid accumulation occurred preferentially in cytosolic LDs. Thus, inhibition of autophagy triggered increased TG and LD accumulation in hepatocytes challenged with a lipid stimulus.

Autophagy is required for LD breakdown

To determine how autophagy regulates TG levels, rates of TG synthesis and FFA β -oxidation were examined in siAtg5 cells. Equivalent increases in TG synthesis occurred in control and siAtg5 cells in response to oleate or culture in MCDM (Fig. 2a). Rates of β -oxidation, indicative of the levels of FFA generated by TG hydrolysis¹⁰, increased during lipid loading, but to a much lesser extent in cells with inhibited autophagy (Fig. 2b), consistent with reduced lipolysis. TG breakdown in cells cultured in oleate or MCDM was significantly decreased in siAtg5 cells (Fig. 2c, d). TG content after treatment with the lipolysis inhibitor diethylumbelliferyl phosphate (DEUP)¹¹ was higher than after inhibition of autophagy with 3MA (Fig. 2e), consistent with a blockage of all lipolysis by DEUP (Supplementary Fig. 6a)^{12,13}, but a partial inhibition by a loss of autophagy. Co-treatment with 3MA did not have an additive effect on the DEUP-mediated increase (Fig. 2e), supporting the theory that lipid accumulation during autophagy inhibition resulted from blocked lipolysis. DEUP did not affect autophagy, because autophagic flow (determined by changes in levels of LC3-II (also known as MAP1LC3B) in the absence and presence of lysosomal inhibitors) was unaffected by DEUP (Supplementary Fig. 6b).

Autophagic pathway components associate with LDs

To confirm that lysosomes regulate intracellular lipid, the effect of inhibiting lysosomal hydrolysis on lipid stores was examined. Lysosomal inhibition increased cellular TG and cholesterol content (Supplementary Fig. 7a, b) as well as LD accumulation (Supplementary Fig. 7c) in the absence or presence of a lipid stimulus. Double immunofluorescence studies revealed increased co-localization of LDs with the lysosome-associated membrane protein type 1 (LAMP1) in oleate or MCDM (Fig. 2f and Supplementary Fig. 8). In the absence of a stimulus for TG formation, LD/LAMP1 co-localization was observed only when lysosomal hydrolysis was inhibited, indicating that rapid LD turnover occurred in lysosomes under these conditions (Supplementary Fig. 8). In contrast to MCDM, where lysosomal inhibition increased LD/LAMP1 co-localization to some extent, this treatment did not modify LD association to lysosomes with oleate supplementation. This finding could represent an inability of the autophagic system to accommodate to the sudden FFA increase. Also supporting this possibility, autophagic flow (measured as LC3-II degradation or increased LC3-positive puncta by immunofluorescence; Supplementary Fig. 9a–c) was twofold greater in cells in MCDM

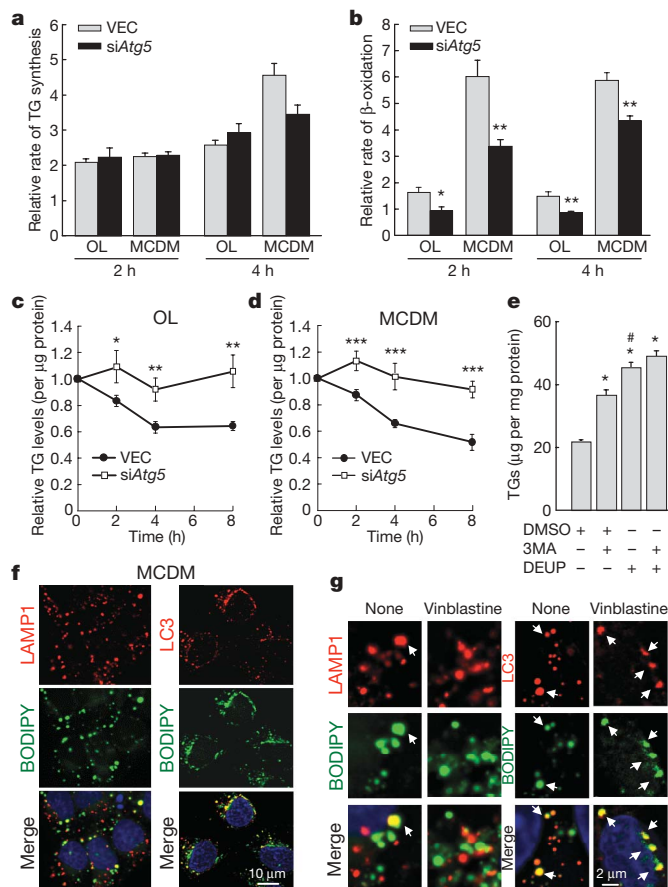


Figure 2 | Inhibition of autophagy decreases TG β -oxidation and decay.

a, b, VEC and siAtg5 cells cultured with oleate (OL) or in MCDM were examined for their rates of TG synthesis (**a**) and β -oxidation (**b**) as compared to cells in regular medium alone (* $P < 0.03$, ** $P < 0.004$ with VEC cells in the same medium, $n = 3-4$). **c, d,** Rates of TG decay in OL (**c**) and MCDM (**d**) (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ as compared to VEC cells, $n = 3-7$). **e,** TG levels in wild-type cells treated with dimethyl sulphoxide vehicle (DMSO), 3-methyladenine (3MA) or diethylumbelliferyl phosphate (DEUP) (* $P < 0.00001$ with DMSO-treated cells, # $P < 0.003$ with 3MA-treated cells, $n = 6$). Error bars, s.e.m. **f,** Co-localization of BODIPY 493/503 (green) with LAMP1 (red) or LC3 (red) in hepatocytes in MCDM. **g,** High-magnification regions of hepatocytes in MCDM alone (none) or treated with vinblastine and stained as labelled. Arrows indicate co-localization.

compared to regular medium, but did not change with oleate. Despite increased autophagic flow, only a moderate decrease in the activity (20–30%) of the autophagy repressor mTOR (also known as FRAP1) as determined by autophosphorylation and phosphorylation of its downstream substrate p70S6K (also known as RPS6KB1) was observed with MCDM (Supplementary Fig. 9d). In addition, the absence of change in the autophagy activator beclin 1 indicates that autophagy was not induced in response to lipid stimuli. Instead, basal autophagy may be primarily responsible for regulating cellular lipid storage. Consistent with this conclusion, induction of autophagy further decreased lipid stores. Treatment with rapamycin or lithium chloride, activators of autophagy, significantly decreased LD number and TGs, and increased LD/LAMP1 co-localization (more evident with lysosomal inhibition) with lipid stimuli (Supplementary Figs 8 and 10a). As for basal autophagy, the flow of rapamycin-induced autophagy (measured as the increase in LD/LAMP1 co-localization with lysosomal proteolysis inhibition) was considerably reduced in cells exposed to oleate (Supplementary Fig. 8). The different autophagy effectors were unaffected by oleate (Supplementary Fig. 9d), indicating that autophagosome formation is preserved but their clearance is compromised to some extent in these cells and suggesting that increased LDs

result from both augmented LD formation and diminished lysosomal breakdown.

In support of autophagy mediating delivery of LD content to lysosomes, LD/LAMP1 co-localization was markedly reduced by inhibitors of autophagosome formation (3MA) or autophagosome–lysosome fusion (vinblastine; Fig. 2g and Supplementary Fig. 8). Similarly, LD/LAMP1 co-localization was lower in siAtg5 cells and did not increase when lysosomal proteolysis was blocked (Supplementary Fig. 11a). Furthermore, LD co-localization with the autophagosome marker LC3 demonstrated a direct association between LDs and autophagosomes (Fig. 2f and Supplementary Fig. 12). In the absence of a stimulus for lipid accumulation, LD/LC3 co-localization was more prominent with lysosomal inhibition (Supplementary Fig. 10b), supporting a constitutive function for autophagy in LD regulation. Induction of autophagy by rapamycin or lithium chloride also increased LD/LC3 co-localization in untreated and oleate-treated cells (Supplementary Figs 10b, c and 12). The lack of a significant increase in LD/LC3 co-localization in cells in MCDM during autophagic induction is probably the consequence of their increased autophagic flux ensuring efficient lysosomal clearance of newly formed LD-containing autophagosomes (Supplementary Fig. 12). In contrast to the blocking effect of vinblastine on LD delivery to lysosomes, this drug did not decrease LD/LC3 co-localization (Fig. 2g and Supplementary Fig. 12). This result indicates that LD engulfment by LC3-positive membranes does not require microtubules, arguing against co-localization representing fusion between LDs and previously formed autophagosomes.

Fluorescence real-time video microscopy revealed that BODIPY 493/503-labelled structures (presumably complete LDs and LD-containing vesicles) and lysosomes associate in a dynamic manner (Supplementary Fig. 13; also see Supplementary Videos 1 and 2). Triple labelling for lysosomes, lipids and TIP47 confirmed that all LD components (lipids and proteins) were delivered to lysosomes (Supplementary Fig. 11b). Electron microscopy was used to elucidate further the mechanism of LD sequestration by autophagic vesicles. LDs were easily identifiable as round light-density structures, not limited by a bilayer lipid membrane (Fig. 3a, inset in the top left panel), with homogenous amorphous content and an average diameter of 0.5 μm that increased 10–15-fold in response to lipid stimuli (Supplementary Fig. 14). Double-membrane structures occupying up to 80% of a single LD were identified (Fig. 3a), along with similar density cytosolic autophagolysosome-like vesicles one-tenth of the size of a LD. These vesicles were surrounded by a double membrane and could have originated from sequestration of a portion of a large LD or an entire small LD (Fig. 3a, top and bottom right panels). Immunogold labelling revealed the presence of LC3 on the LD (often concentrated around membranous structures) and on the smaller lipid-containing double-membrane vesicles (autolipophagosomes; Fig. 3b). In some instances a small dense region heavily labelled for LC3 was present in the proximity of a LD or associated to its surface (Fig. 3b, top left panels). In cells in regular media, only a small percentage of double-membrane vesicles with content of density similar to LDs were detected (<20%), and of these less than half contained only lipid cargo (Fig. 3c). However, both stimuli for lipid accumulation increased the number of lipid-containing vesicles up to 80% with almost half containing only lipid, indicating some level of selectivity towards lipid cargo (Fig. 3c). Prolonged serum removal, a known stimulus of lipolysis, also increased LD/LC3 co-localization in hepatocytes to levels comparable to those from a lipid stimulus (Supplementary Fig. 15a). In contrast, co-localization of LC3 with mitochondria, a common cargo of autophagic vesicles, decreased after prolonged serum removal (Supplementary Fig. 15b). These results indicate that, although basal autophagy of LDs is part of an ‘in bulk’ nonselective process, conditions of increased lipolysis switch selectivity towards macroautophagic engulfment of LDs or macrolipophagy.

Autophagy regulates hepatic lipid stores *in vivo*

Starvation induces hepatic autophagy and increases delivery to the liver of FFAs from adipose tissue lipolysis. Electron microscopy revealed

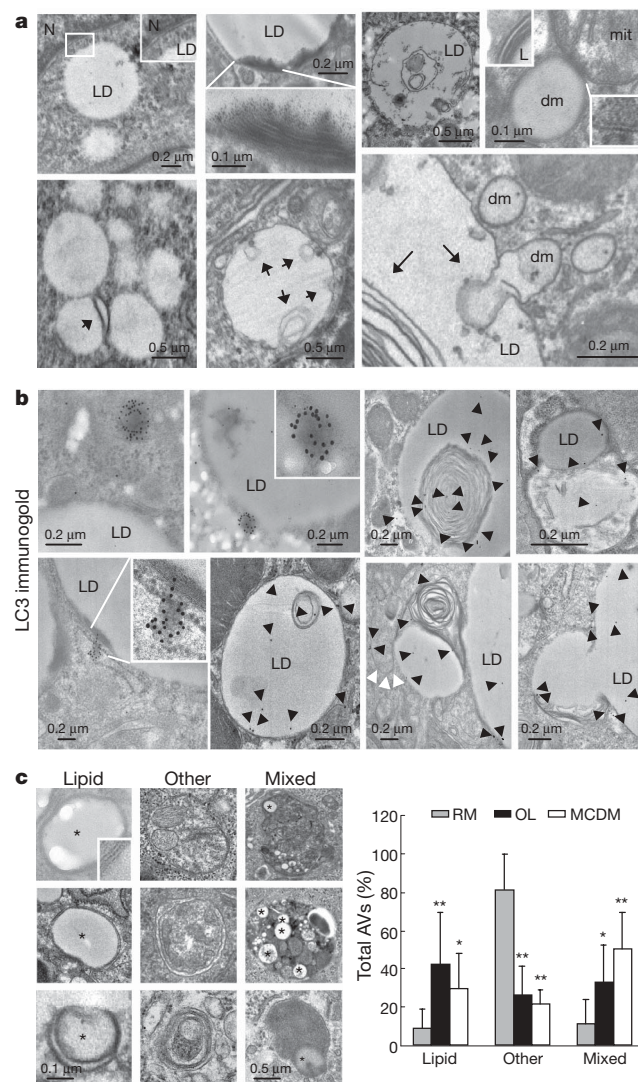


Figure 3 | Lipid droplet content is delivered to lysosomes in autophagosomes. **a**, Electron micrographs of cultured hepatocytes. dm, double-membrane cytosolic vesicles. Arrows indicate membranes in LDs. Insets show the double membrane in nucleus (N), lipid-containing vesicles (L) and mitochondria (mit), but not in LDs. **b**, Mouse liver LC3 immunogold. Insets show higher magnification. Arrowheads indicate gold particles (black) and LC3-labelled bilayer membranes (white). **c**, Percentages of autophagic vacuoles (AVs) containing only lipid (Lipids, *), other cargo (Other) or mixed cargo (Mixed) in cells treated as in **b** (* $P < 0.01$, ** $P < 0.001$ with cells in RM, $n = 4-6$). Left: representative examples. Error bars, s.e.m.

that starvation increased the frequency of LDs with areas of increased density and asymmetrically localized multi-membrane structures (Supplementary Fig. 16a) and the number of lipid-containing autolipophagosomes, autophagolysosomes and lysosomes (Supplementary Fig. 16b). LC3-II was undetectable in LDs isolated from the livers of fed mice, but co-purified with LDs from mice starved for 24 h (Fig. 4a). Relative LD purity was demonstrated by enrichment in LD-associated protein adipocyte differentiation-related protein (ADRP, also known as ADFP) and the absence of cytosolic proteins (Fig. 4a). Concomitantly, the percentage of lipid-containing autophagosomes was higher with starvation in both whole liver sections (Fig. 4b) and isolated fractions enriched in autophagosomes/autophagolysosomes (Supplementary Fig. 16b). During the first 24 h of starvation there was a significant increase in autophagic vacuoles with a mixed content (lipids, organelles and cytosol), whereas with progressive starvation

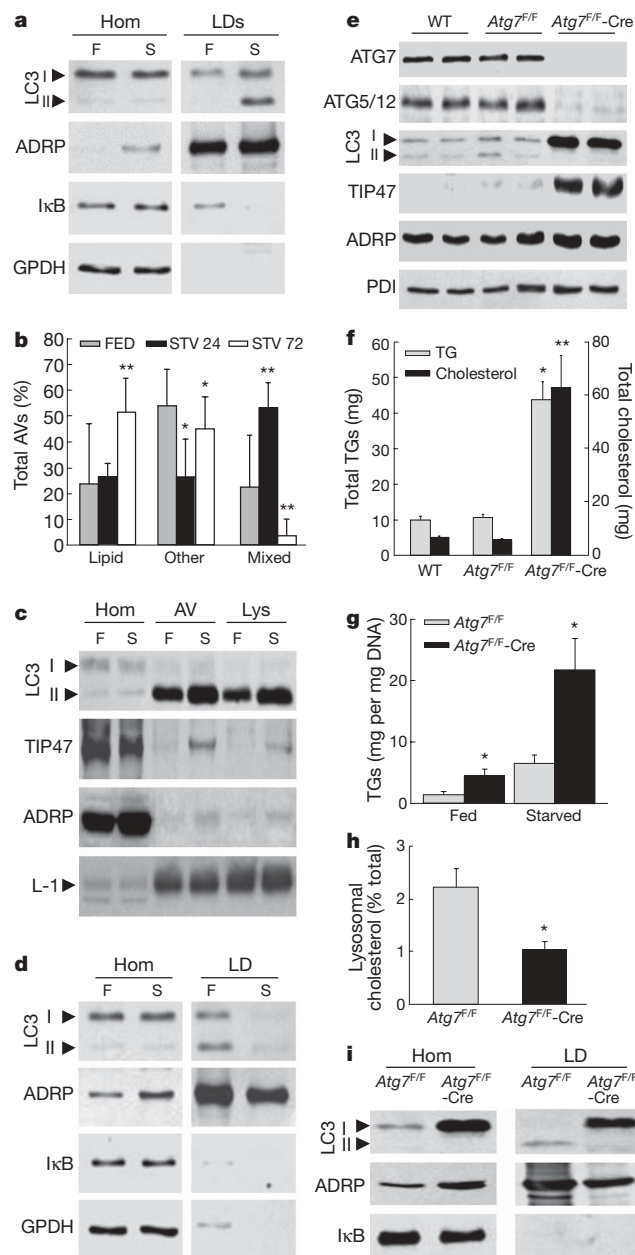


Figure 4 | Effects of starvation, HFD feeding and a hepatocyte-specific blockage of autophagy on hepatic lipid accumulation. **a**, Immunoblots of cellular homogenates (Hom) and LDs from fed (F) or 24-h starved (S) mice. IκB, inhibitor of the nuclear factor of kappa light polypeptide gene enhancer; GPDH, glyceraldehyde-3-phosphate dehydrogenase. **b**, Percentages of autophagic vacuoles (AVs) containing only lipid, other cargo or mixed cargo calculated from samples processed as in Supplementary Fig. 16b ($*P < 0.001$, $**P < 0.0001$, $n = 4-6$). **c**, Immunoblots of Hom, AV and lysosomes (Lys). L-1, LAMP1. **d**, Immunoblots of liver Hom and LDs from HFD-fed mice. **e**, Liver homogenate immunoblots. PDI, protein disulphide isomerase. **f**, Total hepatic TG and cholesterol content ($*P < 0.01$, $**P < 0.00001$, $n = 8-17$). **g**, Hepatic TG concentration ($*P < 0.05$, $n = 3$). **h**, Percentage of cellular cholesterol in lysosomes ($*P < 0.02$, $n = 4$). **i**, Immunoblots of homogenates and isolated LDs. Error bars, s.e.m.

autolipophagosomes become more abundant (Fig. 4b), supporting the selectivity of macrolipophagy. In fact, the amount of LD proteins and lipids in autophagic vacuoles and lysosomes increased with starvation (Fig. 4c and Supplementary Fig. 17a, b) whereas lipid content in the endoplasmic reticulum remained unchanged (Supplementary Fig. 17c, d). The increase in lipids in autophagic compartments and the presence of LC3-II in LDs occurred gradually starting within 6 h of fasting (Supplementary Fig. 17e-g). Thus, induction of

autophagy and hepatic FFA delivery during starvation increased the association of autophagic components to LDs and the presence of LD components as autophagic cargo.

To confirm the relative inability of autophagy to accommodate to an exogenous lipid load observed *in vitro*, the association between LC3 and LD was analysed in mice fed regular diet or 16 weeks of a high-fat diet (HFD). Along with the expected increase in LD number and size (Supplementary Fig. 14b), a marked reduction (from 72% to 37%) in the LD-containing membranous structures or autophagosome-like vesicles was seen with starvation in HFD-fed mouse livers by electron microscopy (Supplementary Fig. 18). Autophagic vacuoles were still observed in starved HFD-fed mice, but the percentage of autolipophagosomes was markedly decreased. Surprisingly, immunogold staining (Supplementary Fig. 18) and immunoblot analysis of isolated LDs (Fig. 4d) revealed that LC3-II still associated with LDs of HFD-fed mice, but this association was more prominent under fed rather than starved conditions. Overall, these results are consistent with those in cultured cells that an exogenous lipid challenge reduces the efficiency of the interaction between the autophagic system and LDs.

For direct evidence of the regulation of hepatic lipids by autophagy *in vivo*, TG content was examined in *Atg7^{F/F}*-Alb-Cre mice with a hepatocyte-specific knockout of the autophagy gene *Atg7*. *Atg7^{F/F}*-Alb-Cre mice had decreased levels of ATG7, conjugated ATG5 and LC3-II in the liver but not in other organs (Fig. 4e and Supplementary Fig. 19a, b), and a marked increase in liver size, as previously reported⁴. At four months of age, oil red O staining of liver sections (Supplementary Fig. 19c), levels of total TGs (Fig. 4f), TGs normalized to DNA content (Supplementary Fig. 19d), LD-associated proteins TIP47 and ADRP (Fig. 4e) and TG accumulation after starvation (Fig. 4g) were increased in *Atg7^{F/F}*-Alb-Cre mouse livers compared to control mice. Hepatic total cholesterol content and concentration were also increased in the knockout animals (Fig. 4f and Supplementary Fig. 19e), but the percentage of cholesterol in lysosomes was significantly decreased (Fig. 4h). TGs secreted as VLDL (also known as APOA5) by the liver, derived predominantly from the hydrolysis of TGs in LDs¹⁴, was also significantly decreased in *Atg7^{F/F}*-Alb-Cre mice (Supplementary Fig. 19f). Isolated LDs from *Atg7^{F/F}*-Alb-Cre mice had no LC3-II but higher levels of LC3-I (Fig. 4i). The fact that LC3-I was still detected in these LDs, as well as in LDs from *Atg5* knockout cells (Supplementary Fig. 10d), indicates that conjugation of LC3 is not required before association to LDs, but occurs on the surface of this organelle.

Discussion

These findings demonstrate a previously unknown interrelationship between autophagy and lipid metabolism that we term macrolipophagy (Supplementary Fig. 1). Autophagy regulates lipid content because: inhibition of autophagy increased TGs and LDs *in vitro* and *in vivo*; loss of autophagy decreased TG breakdown; TGs and LD structural proteins co-localized with autophagic compartments; and LC3 associated with LDs. Moreover, a reverse relationship exists in which an abnormal increase in intracellular lipid impairs autophagic clearance as shown by decreased LD/LAMP1 co-localization and the absence of autophagic upregulation in hepatocytes cultured with lipids, as well as reduced association of autophagic vacuoles with LDs in response to starvation in HFD-fed mice. This interrelationship may trap hepatocytes in a harmful cycle in which decreased autophagy promotes lipid accumulation that then further suppresses autophagic function, thereby additionally increasing lipid retention (Supplementary Fig. 1).

Although future studies must clarify the mechanism of sequestration/degradation of LDs through autophagy, our data indicate that lysosomes do not fuse directly with LDs but with LD-containing autophagosomes. On the basis of morphometric analysis and LC3 immunogold staining we propose that LC3 is recruited to the LDs where LC3 initiates the formation of a limiting membrane through ATG7-dependent conjugation. Some of the lipid cargo observed

within autophagosomes could originate from sequestration of whole small LDs alone or with other cytosolic components that may also become trapped in the sealing vesicle (middle panel, Supplementary Fig. 1). In contrast, for larger LDs, the autophagosome only sequesters a discrete region of the LD, which then pinches off as a double-membrane vesicle enriched in LC3 (autolipophagosome). Such partial sequestration of organelles by autophagy has been described previously for the endoplasmic reticulum and nucleus (piecemeal microautophagy)^{15–17}.

Decreased autophagy in the liver with ageing¹⁸ may contribute to hepatic lipid accumulation that occurs along with an increased incidence of the metabolic syndrome in aged humans¹⁹. The ability of increased lipid content to impair autophagy also indicates that lipid accumulation could contribute to the decrease in autophagic function with ageing. Therapeutic strategies to increase autophagic function may therefore provide a new approach to prevent the metabolic syndrome and its associated pathologies.

METHODS SUMMARY

The hepatocyte cell line RALA255-10G was cultured under nontransformed conditions, as described previously²⁰. Wild-type and *Atg5*^{−/−} mouse embryonic fibroblasts provided by N. Mizushima²¹ were cultured as described previously²². TG content was determined by the Trig/GB Kit (Roche Diagnostics), cholesterol content by the Amplex Red Cholesterol Assay (Invitrogen), fatty acid β -oxidation by a modification of a previously used method²³, and TG decay in cells radio-labelled with [¹⁴C]oleate and TG synthesis by standard methods¹². shRNAs were cloned into pSUPER (Ambion) and then pCCL.sin.PPT.hPGK.GFPwpre²⁴. Protein isolation and western blotting were performed as described previously²⁵. Fluorescence microscopy for BODIPY 493/503 (Invitrogen) and immunofluorescence were performed as described previously²⁶. *Atg7*^{F/F} mice⁴ were crossed with Alb-Cre mice²⁷ to generate *Atg7*^{F/F}-Alb-Cre mice. Some animals were fed a high-fat diet (60% kcal in fat; Research Diets, D12492). Electron microscopy and immunogold labelling were performed as described previously²⁶. LDs from mouse livers were isolated by sucrose density gradient centrifugation²⁸ and autophagic vacuoles and lysosomes by centrifugation in metrizamide discontinuous density gradients²⁹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 November 2008; accepted 12 March 2009.

Published online 1 April 2009.

- Martin, S. & Parton, R. G. Lipid droplets: a unified view of a dynamic organelle. *Nature Rev. Mol. Cell Biol.* **7**, 373–378 (2006).
- Zechner, R., Strauss, J. G., Haemmerle, G., Lass, A. & Zimmermann, R. Lipolysis: pathway under construction. *Curr. Opin. Lipidol.* **16**, 333–340 (2005).
- Finn, P. F. & Dice, J. F. Proteolytic and lipolytic responses to starvation. *Nutrition* **22**, 830–844 (2006).
- Komatsu, M. *et al.* Impairment of starvation-induced and constitutive autophagy in *Atg7*-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
- Mizushima, N., Levine, B., Cuervo, A. M. & Klionsky, D. J. Autophagy fights disease through cellular self-digestion. *Nature* **415**, 1069–1075 (2008).
- Mizushima, N. & Klionsky, D. J. Protein turnover via autophagy: implications for metabolism. *Annu. Rev. Nutr.* **27**, 19–40 (2007).
- Blommaert, E. F., Krause, U., Schellens, J. P., Vreeling-Sindelarova, H. & Meijer, A. J. The phosphatidylinositol 3-kinase inhibitors wortmannin and LY294002 inhibit autophagy in isolated rat hepatocytes. *Eur. J. Biochem.* **243**, 240–246 (1997).
- Leclercq, I. A. *et al.* CYP2E1 and CYP4A as microsomal catalysts of lipid peroxides in murine nonalcoholic steatohepatitis. *J. Clin. Invest.* **105**, 1067–1075 (2000).
- Sahai, A. *et al.* Roles of phosphatidylinositol 3-kinase and osteopontin in steatosis and aminotransferase release by hepatocytes treated with methionine-choline-deficient medium. *Am. J. Physiol. Gastrointest. Liver Physiol.* **291**, G55–G62 (2006).
- Owen, O. E., Reichard, G. A. Jr, Patel, M. S. & Boden, G. Energy metabolism in feasting and fasting. *Adv. Exp. Med. Biol.* **111**, 169–188 (1979).

- Kellner-Weibel, G., McHendry-Rinde, B., Haynes, M. P. & Adelman, S. Evidence that newly synthesized esterified cholesterol is deposited in existing cytoplasmic lipid inclusions. *J. Lipid Res.* **42**, 768–777 (2001).
- Brasaemle, D. L. *et al.* Perilipin A increases triacylglycerol storage by decreasing the rate of triacylglycerol hydrolysis. *J. Biol. Chem.* **275**, 38486–38493 (2000).
- Gocze, P. M. & Freeman, D. A. Factors underlying the variability of lipid droplet fluorescence in MA-10 Leydig tumor cells. *Cytometry* **17**, 151–158 (1994).
- Wiggins, D. & Gibbons, G. F. The lipolysis/esterification cycle of hepatic triacylglycerol. Its role in the secretion of very-low-density lipoprotein and its response to hormones and sulphonylureas. *Biochem. J.* **284**, 457–462 (1992).
- Bernales, S., McDonald, K. L. & Walter, P. Autophagy counterbalances endoplasmic reticulum expansion during the unfolded protein response. *PLoS Biol.* **4**, e423 (2006).
- Roberts, P. *et al.* Piecemeal microautophagy of nucleus in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **14**, 129–141 (2003).
- Yorimitsu, T., Nair, U., Yang, Z. & Klionsky, D. J. Endoplasmic reticulum stress triggers autophagy. *J. Biol. Chem.* **281**, 30299–30304 (2006).
- Cuervo, A. M. *et al.* Autophagy and aging: the importance of maintaining “clean” cells. *Autophagy* **1**, 131–140 (2005).
- Ford, E. S., Giles, W. H. & Dietz, W. H. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *J. Am. Med. Assoc.* **287**, 356–359 (2002).
- Wang, Y., Schattenberg, J. M., Rigoli, R. M., Storz, P. & Czaja, M. J. Hepatocyte resistance to oxidative stress is dependent on protein kinase C-mediated down-regulation of c-Jun/AP-1. *J. Biol. Chem.* **279**, 31089–31097 (2004).
- Mizushima, N. *et al.* Dissection of autophagosome formation using Apg5-deficient mouse embryonic stem cells. *J. Cell Biol.* **152**, 657–668 (2001).
- Wang, Y. *et al.* Loss of macroautophagy promotes or prevents fibroblast apoptosis depending on the death stimulus. *J. Biol. Chem.* **283**, 4766–4777 (2008).
- Hoppel, C., DiMarco, J. P. & Tandler, B. Riboflavin and rat hepatic cell structure and function. Mitochondrial oxidative metabolism in deficiency states. *J. Biol. Chem.* **254**, 4164–4170 (1979).
- Piva, R. *et al.* Ablation of oncogenic ALK is a viable therapeutic approach for anaplastic large-cell lymphomas. *Blood* **107**, 689–697 (2006).
- Schattenberg, J. M., Wang, Y., Singh, R., Rigoli, R. M. & Czaja, M. J. Hepatocyte CYP2E1 overexpression and steatohepatitis lead to impaired hepatic insulin signaling. *J. Biol. Chem.* **280**, 9887–9894 (2005).
- Kaushik, S., Massey, A. C. & Cuervo, A. M. Lysosome membrane lipid microdomains: novel regulators of chaperone-mediated autophagy. *EMBO J.* **25**, 3921–3933 (2006).
- Postic, C. *et al.* Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J. Biol. Chem.* **274**, 305–315 (1999).
- Brasaemle, D. L. & Wolins, N. E. Isolation of lipid droplets from cells by density gradient centrifugation. *Curr. Protoc. Cell Biol.* Chapter 3, unit 3.15 (2006).
- Cuervo, A. M., Palmer, A., Rivett, A. J. & Knecht, E. Degradation of proteasomes by lysosomes in rat liver. *Eur. J. Biochem.* **227**, 792–800 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank D. Silver for his discussions, N. Mizushima for providing the *Atg5*^{−/−} mouse embryonic fibroblasts, R. Stockert for the protein disulphide isomerase antibody and the personnel at the Analytical Imaging Facility for their technical assistance. This work was supported by National Institutes of Health grants from the National Institute of Diabetes and Digestive and Kidney Diseases and National Institute on Aging, a Glenn Award and an American Liver Foundation Postdoctoral Research Fellowship Award (R.S.).

Author Contributions R.S. performed biochemical analyses and immunoblots. S.K. performed the imaging studies and subcellular fractionations. Y.W. generated the shRNAs and performed immunoblotting. Y.X. performed biochemical analyses. R.S., Y.W., Y.X. and I.N. all contributed to the *in vivo* studies. M.K. and K.T. provided the knockout mice. A.M.C. and M.J.C. conceived and planned the study, analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.J.C. (czaja@aecom.yu.edu) or A.M.C. (amcuervo@aecom.yu.edu).

METHODS

Cells and cell culture. The rat hepatocyte line RALA255-10G was cultured under nontransformed conditions, as described previously²⁰. Some cells were cultured in high glucose or serum-free DMEM lacking methionine and choline (Atlanta Biologicals) or pre-treated with 10 mM 3-methyladenine, 100 μ M DEUP, 50 μ M chloroquine (Sigma), 20 mM ammonium chloride or 100 μ M leupeptin (Fisher). Oleic and palmitic acid were conjugated to albumin, as described previously³⁰, and cells were treated with 0.125 or 0.25 mM oleate or palmitate for 24 h in high glucose, serum-free DMEM. Mouse embryonic fibroblasts were maintained in high glucose DMEM with 10% fetal bovine serum.

Lentiviral *Atg5* siRNA construction. The following small hairpin (sense–loop–antisense) RNAs (shRNA) for *Atg5* were cloned into lentiviral vectors: 5'-GA TCCCGCTCAGGTGATCAACGAAATTTCAAGAGAATTCGTTGATCACC TGACTTTTTC-3' (sense), 5'-TCGAGAAAAAGTCAGGTGATCAACGAAAT TCTCTGAAATTCGTTGATCACC TGACGGG-3' (antisense) for sh*Atg5*, and 5'-GATCCCGCTGTTAGTGAGATTTGGTTTCAAGAGAACCAATCT CACTAACATCTTTTTC-3' (sense), 5'-TCGAGAAAAAGATGTTAGTGAG TTTGGTTCTCTTGAACCAATCTCACTAACATCGGG-3' (antisense) for sh*Atg5* number 2. The hairpins were cloned between the BglII–XhoI sites of pSUPER (Ambion), and after SmaI–XhoI digestion the fragments, which included the H1 promoter-shRNA cassette, were subcloned into the EcoRV–XhoI sites of the vector pCCL.sin.PPT.hPGK.GFPWpre²⁴. Lentiviral stocks were prepared by calcium phosphate transfection of these vectors and the packaging vectors pMDLg/pRRE, pRSV-Rev and pMD2.VSVG into HEK-293T cells. Supernatants were collected over 36 to 48 h, titered by plaque assay and used at a multiplicity of infection of 5 to infect RALA hepatocytes. The efficiency of infection at 72 h, determined by the number of green-fluorescent-protein-positive cells, exceeded 98% for all constructs.

Western blotting. Cell lysates and liver homogenates were subjected to western blot analysis, as previously described²⁵. Membranes were incubated with the following primary antibodies: rabbit anti-ATG7, rabbit anti-LC3, rabbit anti-total and phospho-(Ser2448 and Ser2481) mTOR, rabbit anti-total and phospho-p70S6K, rabbit anti-total and phospho-Akt (Cell Signaling Technology), guinea pig anti-ADRP (Progen Biotechnik), rabbit anti-ATG5 (Novus Biologicals), mouse anti-beclin 1 (BD Biosciences), mouse anti-GPDH (Abcam), rat anti-LAMP1 (Developmental Studies Hybridoma Bank, University of Iowa), rabbit anti-IkB (Santa Cruz Biotechnology) and rabbit anti-TIP47 (ProSci Incorporated). Western blot for β -actin (AbCam) or protein disulphide isomerase (a gift from R. Stockert) was used as loading control.

As a measure of autophagic flow, immunoblots for LC3 were performed in untreated cells and cells treated with the lysosomal inhibitors ammonium chloride and leupeptin. Autophagic flow was determined by the ratio of the densitometric value for LC3-II in the presence of inhibitors to that in the absence of inhibitors, as described previously⁵.

Fluorescence microscopy. Cells were fixed with 3% paraformaldehyde, blocked and incubated with the primary and corresponding Cy5- and texas-red-conjugated secondary antibodies. Lipid droplets were stained by incubating cells with BODIPY 493/503 (Invitrogen) for 30 min, fixed and processed for immunofluorescence as described previously²⁶. Lysosomes were highlighted with Lysotracker and mitochondria with Mitotracker (Invitrogen). Mounting medium contained DAPI stain to highlight the cell nucleus. Images were acquired with an Axiovert 200 fluorescence microscope (Carl Zeiss Ltd) with a $\times 63$ objective and 1.4 numerical aperture, subjected to deconvolution with the manufacturer's software and prepared using Adobe Photoshop 6.0 software (Adobe Systems Inc.). Quantification was performed in individual frames after deconvolution and thresholding using ImageJ software (NIH) in a minimum of 20 cells per slide. Particle number was quantified with the 'analyze particles' function in thresholded single sections with size (pixel²) settings from 0.1 to 10 and circularity from 0 to 1. Co-localization was calculated by JACoP plugin in single Z-stack sections of deconvoluted images. Real-time video microscopy was performed using 8-chamber slides in medium buffered with HEPES and maintained at 37 °C in a temperature-controlled stage. Images were captured at 30-s intervals with a $\times 40$ objective and $\times 1.6$ Optovar system, and frames in the different fluorescence channels merged using the manufacturer's software.

Oil red O staining. Oil red O staining was performed, as described previously³¹. Staining was assessed by bright-field microscopy and quantified by the Image J software after appropriate thresholding.

Measurement of TG synthesis. TG synthesis was determined by standard methods¹². Cells were cultured with [¹⁴C]oleate–BSA complex and the cellular lipids extracted with hexane/isopropanol (3:2, vol:vol). Lipids were then dried with nitrogen gas, redissolved into chloroform and resolved by thin-layer chromatography using successive solvent systems containing chloroform, acetone, methanol, acetic acid and water in volumetric ratios of 10:4:2:2:1, and hexane, methanol and

acetic acid in ratios of 80:20:1. Phosphorimages were obtained with a Storm Gel and Blot Imaging System (GE Healthcare).

Fatty acid β -oxidation assay. Rates of fatty acid β -oxidation were determined by a modification of a previously used method²³, in which the rate of carbon dioxide production from the oxidation of [¹⁴C]oleate was measured. Cells were cultured in the presence of [¹⁴C]oleate–BSA complex and the released [¹⁴C]carbon dioxide trapped for 1 h at 37 °C onto filter paper soaked in 100 mM sodium hydroxide. The rate of β -oxidation was calculated as the amount of trapped [¹⁴C]carbon dioxide in relative units produced per mg protein per hour.

TG decay determination. Cells were cultured in the presence of [¹⁴C]oleate–BSA complex for 24 h following which the cells were washed and the medium replaced. At different times TGs were extracted from the cells and quantified by thin-layer chromatography as described previously.

Animals. *Atg7*^{F/F} mice⁴ were crossed with Alb-Cre mice²⁷ to generate *Atg7*^{F/F}-Alb-Cre mice with a hepatocyte-specific knockout of *Atg7*. Starved mice were allowed free access to water. Some animals were fed a HFD (34 g per 100 g diet fat, 0.03 g per 100 g diet cholesterol with 60% of the total kcal in fat; Research Diets, D12492). The HFD was begun at 3 weeks of age and continued for a total of 16 weeks. All studies were approved by the Animal Care and Use Committee of the Albert Einstein College of Medicine and followed the National Institutes of Health guidelines for animal care.

Electron microscopy and immunogold. Cells cultured in monolayers and liver blocks (1 mm³) were fixed in 2.5% glutaraldehyde in 100 mM sodium cacodylate, pH 7.43, and post-fixed in 1% osmium tetroxide in sodium cacodylate followed by 1% uranyl acetate. After ethanol dehydration and embedding in LX112 resin (LADD Research Industries), ultrathin sections were stained with uranyl acetate followed by lead citrate. Immunogold labelling was performed in ultrathin sections of samples fixed in 4% paraformaldehyde/0.1% glutaraldehyde in sodium cacodylate, dehydrated and embedded in Lowicryl. Grids were washed in 50 mM glycine in phosphate buffered saline, blocked, incubated with LC3 antibody for 2 h, washed extensively and incubated with the gold-conjugated secondary antibody (1:100) for 2 h. Control grids were incubated with the secondary antibody alone or with an irrelevant immunoglobulin G. After extensive washing, samples were fixed a second time for 5 min in 2% glutaraldehyde, washed and negatively stained with 1% uranyl acetate for 15 min. All grids were viewed on a JEOL 100CX II transmission electron microscope at 80 kV. Morphometric analysis was performed using ImageJ in 15–20 different micrographs for each condition after thresholding. Autophagic vacuoles were identified using previously established criteria^{32,33}. Autophagic vacuoles (vesicles <0.5 μ m) were classified as autophagosomes when they met two or more of the following criteria: double membranes (complete or at least partially visible), absence of ribosomes attached to the cytosolic side of the membrane, luminal density similar to cytosol, and identifiable organelles or regions of organelles in their lumen. Vesicles of similar size but with a single membrane (or less than 40% of the membrane visible as double), luminal density lower than the surrounding cytosol or multiple single membrane-limited vesicles containing light or dense amorphous material were classified as autophagolysosomes. For the classification of autophagic contents, autophagosomes and autophagolysosomes were grouped under the term autophagic vacuoles. Autophagic vacuoles containing only lipids were those with double membranes, homogenous density comparable to that of LDs and lacking other content.

Isolation of LDs. LDs were isolated from mouse livers by density gradient centrifugation following a modification of a method described previously²⁸. Livers homogenized in 0.25 M sucrose were centrifuged at 6,800g for 5 min at 4 °C. The supernatant and fatty layer were centrifuged at 17,000g for 10 min at 4 °C to pellet autophagic vacuoles and lysosomes. The supernatant was adjusted to 20% sucrose in hypotonic lysis buffer (20 mM Tris, 1 mM EDTA, pH 7.4), placed in a centrifuge tube and sequentially overlaid with 5 ml each of 5% sucrose (w/v) in hypotonic lysis buffer. After centrifugation at 28,000g for 30 min at 4 °C, the LD fraction was collected from the top of the tube. LD fractions were delipidated with acetone followed by sequential washes in acetone/ether (1:1, v:v) and ether. The protein pellets were solubilized in SDS and analysed by western blotting.

Autophagic vacuole isolation. Autophagic vacuoles and lysosomes were isolated from mouse livers by differential centrifugation and centrifugation in discontinuous density gradients of metrizamide, following a protocol modified from ref. 34, as described previously²⁹. A fraction enriched in endoplasmic reticulum resealed vesicles (microsomes) was prepared by centrifugation of the supernatant of the 17,000g centrifugation at 100,000g for 1 h.

Assay of hepatic TG secretion. TG secretion was determined as described previously³⁵. Mice were fasted overnight and injected with 1 g per kg of P-407 (Sigma). TG levels were determined on serum drawn immediately before and 6 h after injection. TG production was calculated from the difference in serum TG levels over the 6 h period and expressed as mg per kg per h.

Statistical analysis. All numerical results are reported as mean and s.e.m., and represent data from a minimum of three independent experiments unless otherwise stated. We determined the statistical significance of the difference between experimental groups in instances of single comparisons by the two-tailed unpaired Student's *t*-test of the means with Sigma Plot (Jandel Scientific) software. In instances of multiple means comparisons, we used one-way analysis of variance (ANOVA) followed by the Bonferroni *post hoc* test to determine statistical significance.

30. Goldstein, J. L., Basu, S. K. & Brown, M. S. Receptor-mediated endocytosis of low-density lipoprotein in cultured cells. *Methods Enzymol.* **98**, 241–260 (1983).
31. Andersson, L. et al. PLD1 and ERK2 regulate cytosolic lipid droplet formation. *J. Cell Sci.* **119**, 2246–2257 (2006).
32. Dunn, W. A. Jr. Studies on the mechanisms of autophagy: maturation of the autophagic vacuole. *J. Cell Biol.* **110**, 1935–1945 (1990).
33. Nixon, R. A. et al. Extensive involvement of autophagy in Alzheimer disease: an immuno-electron microscopy study. *J. Neuropathol. Exp. Neurol.* **64**, 113–122 (2005).
34. Marzella, L., Ahlberg, J. & Glaumann, H. Isolation of autophagic vacuoles from rat liver: morphological and biochemical characterization. *J. Cell Biol.* **93**, 144–154 (1982).
35. Millar, J. S., Cromley, D. A., McCoy, M. G., Rader, D. J. & Billheimer, J. T. Determining hepatic triglyceride production in mice: comparison of poloxamer 407 with Triton WR-1339. *J. Lipid Res.* **46**, 2023–2028 (2005).

ARTICLES

Structural basis for leucine-rich nuclear export signal recognition by CRM1

Xiuhua Dong¹, Anindita Biswas¹, Katherine E. Süel¹, Laurie K. Jackson¹, Rita Martinez¹, Hongmei Gu¹ & Yuh Min Chook¹

CRM1 (also known as XPO1 and exportin 1) mediates nuclear export of hundreds of proteins through the recognition of the leucine-rich nuclear export signal (LR-NES). Here we present the 2.9 Å structure of CRM1 bound to snurportin 1 (SNUPN). Snurportin 1 binds CRM1 in a bipartite manner by means of an amino-terminal LR-NES and its nucleotide-binding domain. The LR-NES is a combined α -helical-extended structure that occupies a hydrophobic groove between two CRM1 outer helices. The LR-NES interface explains the consensus hydrophobic pattern, preference for intervening electronegative residues and inhibition by leptomycin B. The second nuclear export signal epitope is a basic surface on the snurportin 1 nucleotide-binding domain, which binds an acidic patch on CRM1 adjacent to the LR-NES site. Multipartite recognition of individually weak nuclear export signal epitopes may be common to CRM1 substrates, enhancing CRM1 binding beyond the generally low affinity LR-NES. Similar energetic construction is also used in multipartite nuclear localization signals to provide broad substrate specificity and rapid evolution in nuclear transport.

Nuclear localization and export signals (NLSs and NESs) are recognized by karyopherin- β (Kap- β , also known as importin and exportin) proteins to direct macromolecules in and out of the nucleus^{1–4}. Although several classes of NLSs have been characterized^{5–8}, only one class of NES, the LR-NES, is known at present^{9,10}. The 10–15-residue LR-NESs were first identified in HIV-1 Rev and cyclic-AMP-dependent protein kinase inhibitor, and are recognized by the export-Kap- β CRM1 (refs 9–16). Several mutagenesis and computational studies have led to a consensus sequence ϕ -X(2–3)- ϕ -X(2–3)- ϕ -X- ϕ (in which ϕ is Lys, Val, Ile, Phe or Met, X is any amino acid, and the numbers in parentheses denote the number of repeats) that matches most known LR-NESs, but it is so broad that it is found in most helix-containing proteins^{17–21}. Because many proteins with this motif do not undergo nuclear export, other factors in the signal must affect its availability. Like the multipartite PY-NLS^{8,22} and other vague organelle-targeting signals^{23,24}, the LR-NES may be a complex and diverse signal that needs to be described not just by the consensus sequence but also in the structural and physical context of its interactions with CRM1.

Antifungal antibiotic leptomycin B inhibits nuclear export by alkylating Cys 528 of human CRM1 (ref. 25). It is unclear if leptomycin B sterically blocks substrate binding or if it inhibits substrate binding by a conformational change. Effective inhibition of CRM1 by leptomycin B coupled with the availability of a LR-NES sequence motif have facilitated identification of hundreds of broadly functioning CRM1 substrates^{26,27}. One such export substrate is snurportin 1 (SNUPN), which imports spliceosomal U small nuclear ribonucleoprotein particles (snRNPs) into the nucleus²⁸. SNUPN contains an N-terminal importin- β (karyopherin- β)-binding domain (which is known as sIBB) and a carboxy-terminal nucleotide-binding domain (NBD)²⁹. The latter binds the 5'-2,2,7-terminal trimethylguanosine (m₃G) cap of small nuclear RNAs (snRNAs) and the sIBB domain binds karyopherin- β 1 (Kap- β 1, also known as KPNB1 and importin- β) thus importing snRNP particles into the nucleus. CRM1 recycles nuclear SNUPN back to the cytoplasm²⁸. Like other export Kap- β s, CRM1 binds SNUPN and RanGTP cooperatively in the nucleus, and the CRM1–SNUPN–Ran complex is translocated through the

nuclear pore complex. This ternary export complex probably forms through a stable CRM1–SNUPN binary intermediate²⁸. In the absence of atomic structures of CRM1–substrate complexes, the location of substrate-binding sites on CRM1, the physical nature of NES and its mechanism of recognition remain unknown.

Here we present the crystal structure of the CRM1–SNUPN complex. CRM1 is a ring-shaped protein and SNUPN binds the central convex surface of CRM1 by its N-terminal LR-NES and NBD. The LR-NES is a combined α -helical-extended peptide that occupies a hydrophobic groove between the outer helices of CRM1 HEAT repeats 11 and 12. The LR-NES interface explains the hydrophobic consensus and prevalence of intervening electronegative residues. The location of CRM1 Cys 528 in the hydrophobic groove indicates that leptomycin B inhibits substrate binding by the occupation of the LR-NES-binding site. The second NES epitope of SNUPN is the positively charged NBD surface that binds an acidic patch adjacent to the LR-NES groove. Multipartite recognition as observed in SNUPN may be used more generally to increase CRM1 binding beyond the usually low affinity LR-NES.

Overall structure of the CRM1–SNUPN complex

The 2.9 Å CRM1–SNUPN crystal structure was solved by multiwavelength anomalous dispersion (MAD) using native full-length human CRM1 (residues 1–1071) and residues 1–342 of selenomethionine-labelled human SNUPN (Fig. 1a–c and Supplementary Table 1). The last 30 residues of SNUPN are dispensable for CRM1 binding (Supplementary Fig. 1)²⁸ and SNUPN is exported in a CRM1–SNUPN–Ran complex that probably assembles through a stable CRM1–SNUPN intermediate (dissociation constant (K_d) 1.4 μ M; Supplementary Fig. 2)²⁸. The final CRM1–SNUPN model has an R_{free} of 0.269 and an R_{factor} of 0.239 (Supplementary Table 1).

CRM1 has 20 HEAT repeats (H1–H20; Fig. 1a, b). A HEAT repeat generally contains two antiparallel helices A and B, each lining the convex and concave side of the protein. Most of repeat H1, several loops within H1–H3, H8–H9 and the last ten residues of CRM1 are not modelled owing to weak electron density. Repeats H1–H20 are

¹Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas, 6001 Forest Park, Dallas, Texas 75390-9041, USA.

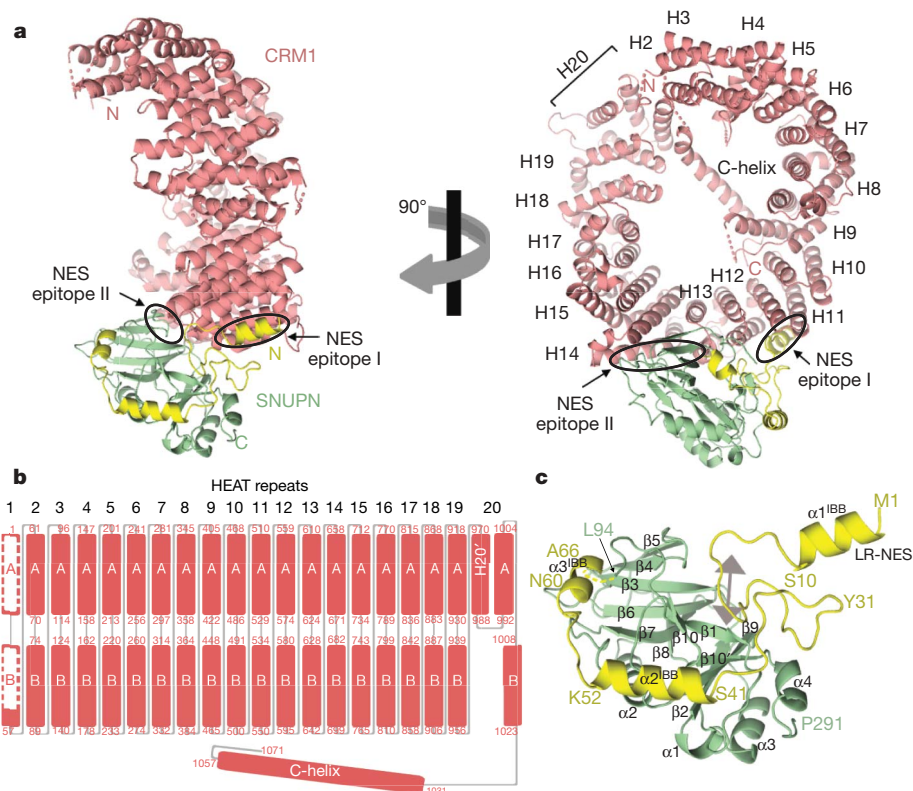


Figure 1 | Overall structure of the CRM1-SNUPN complex. **a**, Orthogonal views of the CRM1-SNUPN complex. CRM1 is pink and the sIBB and NBD domains of SNUPN are yellow and green, respectively. HEAT repeats 2–20 of CRM1 are labelled H2–H20. **b**, HEAT repeat organization of CRM1. Most

arranged in a ring with H2–H5 located within 10 Å of H20. The outer and inner diameters of the CRM1 ring are approximately 90 Å and 45 Å, respectively. H1–H19 are typical two-helical HEAT repeats but H20 has three helices, H20', H20A and H20B. Helix H20B is followed by a 27-residue helix named the C-terminal or C-helix that protrudes across the plane of the ring such that its C terminus contacts B helices of H8–H12 (Fig. 1a, b). Repeats H15–H20 adopt a similar conformation as in a C-terminal fragment of CRM1 (Protein Data Bank (PDB) accession 1W9C; C α root mean squared deviation (r.m.s.d.) 1.57 Å), but the C-helix was not observed in that structure³⁰.

Our SNUPN model includes the sIBB domain (residues 1–66) and NBD (residues 94–293). The CRM1-bound, nucleotide-free NBD is virtually identical to the isolated m₃G-cap-bound domain (PDB accession 1XK5; C α r.m.s.d. 0.62 Å)³¹. In contrast, CRM1-bound sIBB is quite different from Kap- β -bound sIBB^{32,33}. Residues 1–10 of the sIBB are helical and connected by a long loop to α -helices α 2^{IBB} (residues 41–52) and α 3^{IBB} (residues 60–66) (Fig. 1c). Residues 1–17 protrude away from the protein to bind CRM1, whereas the rest of the sIBB drapes onto one face of the NBD (sIBB–NBD interaction buries 2,909 Å² surface area).

SNUPN binds in a bipartite manner to the outer surface of CRM1 that spans H11–H14. Residues 1–16, which form NES epitope I of SNUPN, interact with H11–H12 of CRM1 (Fig. 1a). The second binding element or NES epitope II, mostly loops surrounding the m₃G-cap-binding site, binds outer helices of CRM1 H12–H14. There are no direct contacts between the two NES epitopes, and their interactions with CRM1 bury an area of 3,551 Å². Binding of SNUPN to the outer surface of CRM1 is notable as known Kap- β -substrate interactions, including that of export Kap- β CSE1 (also known as CSE1L) (structural comparison in ref. 33), involve inner surfaces of the karyopherins³⁴. Although Ran increases the CRM1–SNUPN affinity, mutagenesis data described later and structural comparison with the CSE1–Kap60–Ran (Kap60 is also known as Srp1 and

of H1 is disordered and not modelled in the structure. **c**, Structure of CRM1-bound SNUPN. The linker between sIBB and NBD of SNUPN is shown as a dashed yellow line and a grey arrow marks the location of the nucleotide-binding site of this nucleotide-free SNUPN.

KPNA1) export complex suggest that the binary and ternary CRM1 structures are similar^{33,35}.

NES epitope I of snurportin 1 is a classical LR-NES

It was previously reported that SNUPN does not have a LR-NES, and binds CRM1 through residues 1–300 (sIBB and NBD)²⁸. Structural examination of the CRM1–SNUPN intermediate and sequence analysis of SNUPN show that residues 1–16 (NES epitope I) contain a previously undetected LR-NES. Residues 4–14 of segment 1-MEELSQLASSFSVSQ-16 match the ϕ -X(1–3)- ϕ -X(2–3)- ϕ -X- ϕ LR-NES consensus.

Residues 1–10 of SNUPN form an amphipathic α -helix, residues 11–14 adopt an extended conformation and beyond that, residues 15–40 form a long loop (Figs 1c, 2a and Supplementary Fig. 3). The LR-NES (residues 1–16) is a structurally independent module that does not contact the NBD and protrudes away from the rest of SNUPN. Its N-terminal location that immediately precedes a long loop makes it highly accessible (Fig. 1c). Although no previous structure of this SNUPN segment is available, secondary structure prediction programs suggest that residues 2–9 are helical^{36,37}. Discovery of the LR-NES in SNUPN is further supported by complementary biochemical and cellular localization studies (see later).

Recognition of the leucine-rich NES by CRM1

The LR-NES of SNUPN binds a hydrophobic groove between CRM1 helices H11A and H12A (buried surface area 1,770 Å²; Fig. 2 and Supplementary Fig. 3). Helices H11A and H12A are spaced farther apart (average distance between equivalent C α s is 14.5 Å) than other neighbouring A/B helix pairs in CRM1 (average distance between equivalent C α s is 11.5 Å), resulting in a prominent groove. This groove is wide at the N-terminal end of the A helices, gradually narrows to a bottleneck at the third helical turn of the LR-NES and continues as a narrow groove to its C-terminal end, which is capped

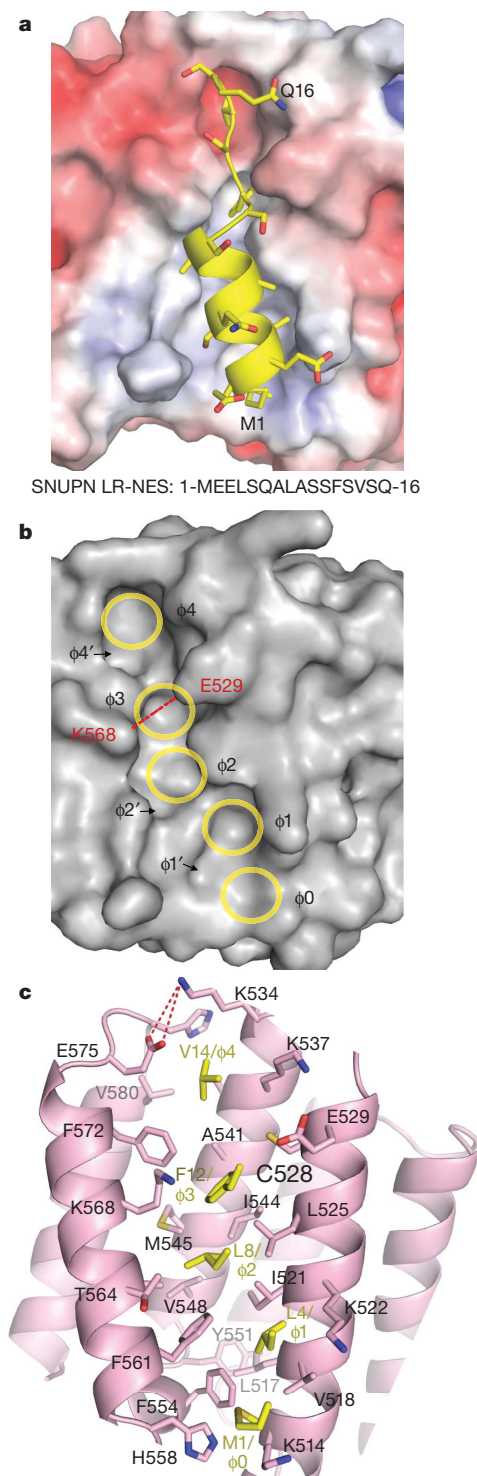


Figure 2 | The LR-NES-binding site. **a**, The LR-NES interface showing the electrostatic surface potential of CRM1 and a ribbon diagram of the SNUPN LR-NES. The amino-acid sequence of the LR-NES of SNUPN is shown. **b**, Molecular surface of the LR-NES binding groove. Individual pockets that bind the five hydrophobic residues of the SNUPN LR-NES are outlined in yellow. Additional pockets are labelled and a dashed line is drawn between Lys 568 and Glu 529 at the narrowest point of the groove. **c**, Residues in the hydrophobic groove of CRM1 (pink) and hydrophobic residues of the SNUPN LR-NES (yellow).

by electrostatic interactions between Lys 534 and Glu 575 (Fig. 2a–c). The highly conserved groove is lined with pockets formed by hydrophobic side chains from helices H11A, H11B and H12A (Fig. 2b, c and Supplementary Fig. 4). Comparison with structures of unliganded

CRM1 and other CRM1–LR-NES complexes will inform on structural flexibility of the groove.

SNUPN side chains Met 1, Leu 4, Leu 8, Phe 12 and Val 14 line one side of the LR-NES and are buried in the LR-NES binding groove by hydrophobic contacts with CRM1 (Fig. 2c and Supplementary Figs 3 and 5). The solvent accessible face of the LR-NES helix is composed of polar residues Glu 2, Glu 3, Ser 5, Gln 6 and Ser 10. Acidic NES side chains Glu 2 and Glu 3 make electrostatic contacts with basic CRM1 side chains Lys 560 and Lys 522 that flank the hydrophobic groove (Fig. 2a, c). Groove-flanking charged side chains on H11A and H12A also contact sIBB residues Ala 22, Tyr 31, Lys 34 and Tyr 25.

We examined interactions between the SNUPN LR-NES and CRM1 using *in vitro* pull-down assays, and compared them to interactions of previously identified LR-NESs. Deletion of the LR-NES helix in glutathione S-transferase (GST)–SNUPN Δ (1–12) or double mutations of Leu 4 and Leu 8 to alanines abolished CRM1–SNUPN interactions, confirming the importance of these hydrophobic NES residues for CRM1 recognition (Fig. 3a). Mutation of electronegative residues Glu 2, Glu 3 and Ser 11 in the NES helix also decreased CRM1 interaction, suggesting a role for polar contacts at this interface.

The isolated SNUPN LR-NES bound CRM1 in the presence of RanGTP, demonstrating sufficiency of the LR-NES in CRM1 binding. The extent and pattern of SNUPN LR-NES binding was comparable to those of LR-NESs from HIV-1 Rev and NMD3 (sequences shown in Supplementary Fig. 6), indicative of their relatively low affinity for CRM1 (Fig. 3b)^{10,38,39}. Complementary LR-NES-binding site mutations of CRM1 quadruple mutant I521A/L525A/F561A/F572A abolished binding to all three LR-NESs, confirming location of the LR-NES-binding site and importance of hydrophobic side

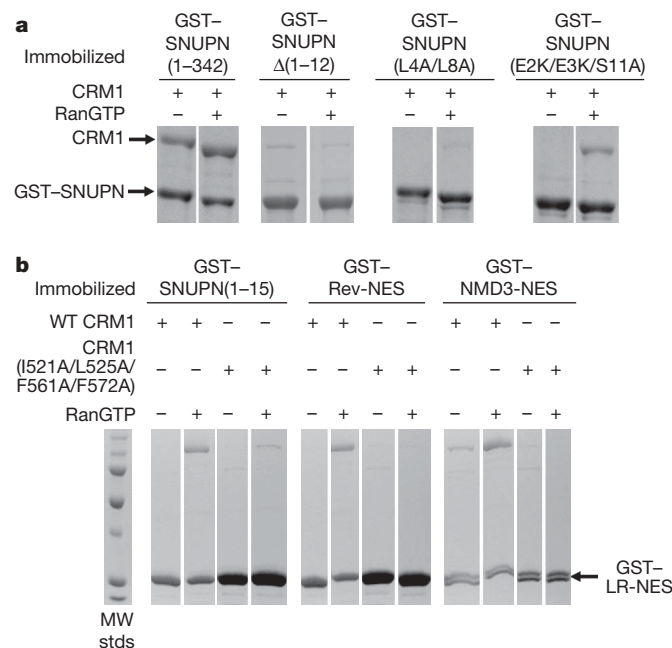


Figure 3 | Hydrophobic residues are critical for LR-NES recognition.

a, Mutations in the LR-NES of SNUPN that affected CRM1 binding. Deletion of the LR-NES (SNUPN Δ (1–12)) and mutation of residues Leu 4 and Leu 8 (SNUPN(L4A/L8A)) abolished interaction with CRM1. Mutation of electronegative residues of the LR-NES (SNUPN(E2K/E3K/S11A)) also decreased CRM1 binding. Bound proteins in the *in vitro* pull-down assays of immobilized GST–SNUPN proteins and CRM1 were visualized by SDS–PAGE and Coomassie staining. **b**, Mutations of hydrophobic residues on CRM1 helices H11A and H12A disrupt interaction with LR-NESs. Immobilized GST–NESs of SNUPN, HIV-1 Rev and NMD3 were incubated with either wild-type (WT) CRM1 or mutant CRM1(I521A/L525A/F561A/F572A) in the presence and absence of RanGTP. Bound proteins were resolved with SDS–PAGE and Coomassie staining. MW stds, molecular weight standards.

chains in the groove. This mutant also confirmed that LR-NESs bind the same hydrophobic groove in the presence and absence of RanGTP. Thus, the binary CRM1–SNUPN complex informs on interactions in the ternary CRM1–SNUPN–Ran complex.

The LR-NES of SNUPN is also a functional export signal in cells. In nuclear transport competition assays in *Saccharomyces cerevisiae*¹², reporter protein SV40 NLS–GFP–SNUPN(1–15) was cytoplasmic even though SV40 NLS–GFP–GFP was nuclear. Mutation of SNUPN Leu 4 and Leu 8 to alanines abolished nuclear export in both the isolated LR-NES and SNUPN (Fig. 4a). Thus, SNUPN(1–15) is a sufficient NES and Leu 4 and Leu 8 are necessary for efficient nuclear export of SNUPN. Furthermore, expression of NLS–GFP–SNUPN(1–15) in *xpo1-1* (or yeast CRM1) temperature-sensitive cells¹² mislocalized the reporter to the nucleus at non-permissive temperature (Fig. 4b), confirming that nuclear export of the SNUPN LR-NES was mediated by CRM1.

Structural and sequence features of a leucine-rich NES

Available structures and secondary structure prediction of LR-NESs suggest that most LR-NESs contain helices¹⁹. A notable structural feature of the SNUPN LR-NES is its helix to extended structural transition, which reflects the hydrophobic pattern of the consensus (Fig. 2a). We labelled hydrophobic positions in the SNUPN LR-NES as $\phi 0$, $\phi 1$, $\phi 2$, $\phi 3$ and $\phi 4$ (consensus $\phi 1$ -X(2–3)- $\phi 2$ -X(2–3)- $\phi 3$ -X- $\phi 4$, in which $\phi 0$ is Met 1; Fig. 2b). Hydrophobic pockets in the CRM1 groove that bind these LR-NES positions are similarly labelled. Positions $\phi 0$ –2 (Met 1, Leu 4 and Leu 8) are on the same side of three consecutive α -helical turns, whereas $\phi 3$ and $\phi 4$ (Phe 12 and Val 14) have the alternating pattern characteristic of a β strand (Fig. 2a–c).

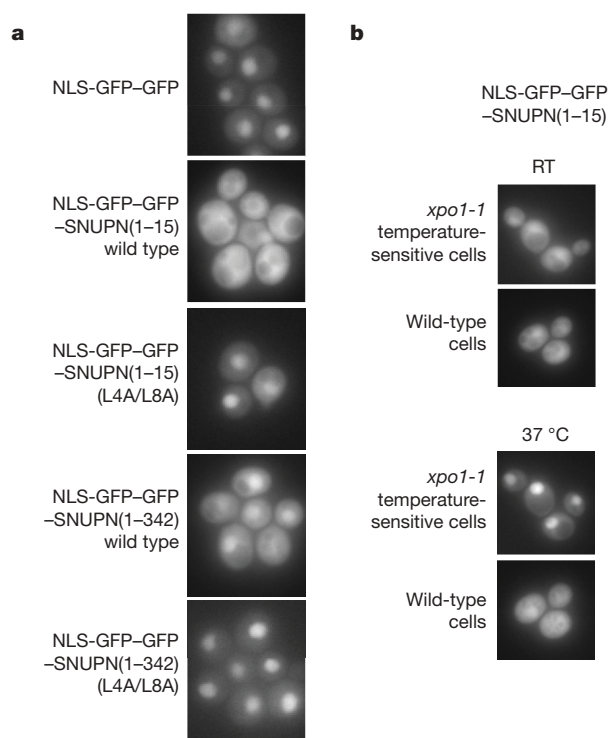


Figure 4 | Localization of the SNUPN and its LR-NES in yeast. **a**, The NLS-GFP–GFP–SNUPN reporter is localized to the cytoplasm in the wild-type *S. cerevisiae* cells, but the NLS-GFP–GFP–SNUPN(L4A/L8A) mutant accumulates in the nucleus. Each panel displays the same GFP fluorescence scale. **b**, Location of NLS-GFP–GFP–SNUPN-NES in *xpo1-1* temperature-sensitive (top) and wild-type (bottom) cells at the permissive (room temperature, RT) and non-permissive (37 °C) temperature. The reporter accumulates in the nucleus of *xpo1-1* temperature-sensitive cells at the non-permissive temperature, but remains cytoplasmic in the wild-type cells.

Several structural features of the CRM1 hydrophobic groove seem to select for the combined helix-extended LR-NES conformation. First, the hydrophobic groove is wide at one end, allowing it to accommodate the LR-NES helix (Fig. 2a, b). Hydrophobic pockets here are appropriately spaced (5.4–6.0 Å) to interact with hydrophobic residues on one side of an α -helix (Fig. 2b). Second, at the third LR-NES helical turn, long CRM1 side chains Glu 529 and Lys 568 constrict the groove, necessitating a structural transition of the LR-NES to an extended conformation (Fig. 2b, c). Unsatisfied helical carbonyls at the break, backbone carbonyl and amide of the extended segment make polar contacts with CRM1 side chains Glu 529, Lys 537 and Lys 568 (Supplementary Fig. 7). Third, CRM1 pockets at the narrow part of the groove are spaced ~ 8 Å apart, matching the distance between side chains on one side of a β strand.

The CRM1 groove contains three other pockets ($\phi 1'$, $\phi 2'$ and $\phi 4'$) adjacent to the $\phi 0$, $\phi 1$, $\phi 2$ and $\phi 4$ pockets (Fig. 2b). These pockets may provide further binding sites to accommodate the variety of hydrophobic side chains and hydrophobic register observed in experimentally identified LR-NESs^{19,40}. Location of the $\phi 1'$ pocket between the $\phi 0$ and $\phi 1$ pockets, and the width of the groove here may also accommodate different LR-NES backbone conformations such as bent and 3_{10} helices, as well as the rare extended conformations, explaining the degeneracy observed at the N-terminal of the consensus¹⁹.

A definite preference for glutamate, aspartate and serine residues has been observed in LR-NES positions not occupied by hydrophobic residues^{19,40}. One side of the LR-NES is exposed to solvent, explaining the presence of polar residues between the hydrophobic positions, whereas the basic surface that flanks the N-terminal half of the CRM1 groove explains the preference for acidic and electronegative residues (Fig. 2a). Previous sequence and structural analyses of LR-NES-containing proteins suggested that the signal needs to be accessible and flexible¹⁹. The N-terminal location of the LR-NES and its connection by a long loop to the rest of SNUPN makes it highly accessible. It will be critical to consider accessibility and structural context of LR-NESs within whole proteins in future efforts to identify the signal. LR-NESs at the termini of proteins and those flanked by long loops are more likely to be true export signals that can be exported by CRM1.

The leptomycin B modification site

Leptomycin B specifically inhibits CRM1 nuclear export^{13–15,27}. The electrophilic α,β -unsaturated δ -lactone of leptomycin B covalently binds the nucleophilic sulphhydryl group of Cys 528 in CRM1 by a Michael-type addition²⁵. Cys 528 is located in the LR-NES binding groove (Fig. 2c) and makes van der Waals contact with SNUPN Phe 12 (Supplementary Fig. 5). Therefore, covalent addition of leptomycin B to the sulphhydryl of Cys 528 will block access to the LR-NES-binding site as part or the entire drug occupies the groove. Leptomycin B prevents formation of the ternary CRM1–SNUPN–Ran complex⁴¹. We show in Supplementary Fig. 8 that leptomycin B also prevents binary CRM1–SNUPN interaction in the absence of Ran, suggesting that the mode of LR-NES binding and inhibition is similar in the presence and absence of Ran.

The C-terminal NES epitope II of SNUPN

The interface between the NBD of SNUPN and H12–H14 of CRM1 is adjacent to the LR-NES binding groove and constitutes the second NES site of the bipartite complex (buried surface area 1,490 Å²; Fig. 5a–c). Many NBD residues at this interface are located near the empty m_3 G-cap-binding site. These include 126–VGK–128 of the $\beta 2$ – $\beta 3$ loop, 143–TKSGYCVN–150 of the $\beta 4$ – $\beta 5$ loop, and Arg 278 of the $\beta 10$ – $\alpha 4$ loop, which surround the nucleotide-binding site, as well as residues 178–EVNQ–181 and 221–KTKLNPF–227 on nearby $\beta 6$ – $\beta 7$ and $\alpha 2$ – $\beta 8$ loops, respectively (Fig. 5b). NES epitope II of SNUPN interacts with CRM1 helices H13A, H14A, H14B and intra-HEAT loops of H12 and H13. Mutations of interface residues on either protein decreased CRM1–SNUPN interactions (Supplementary Fig. 9).

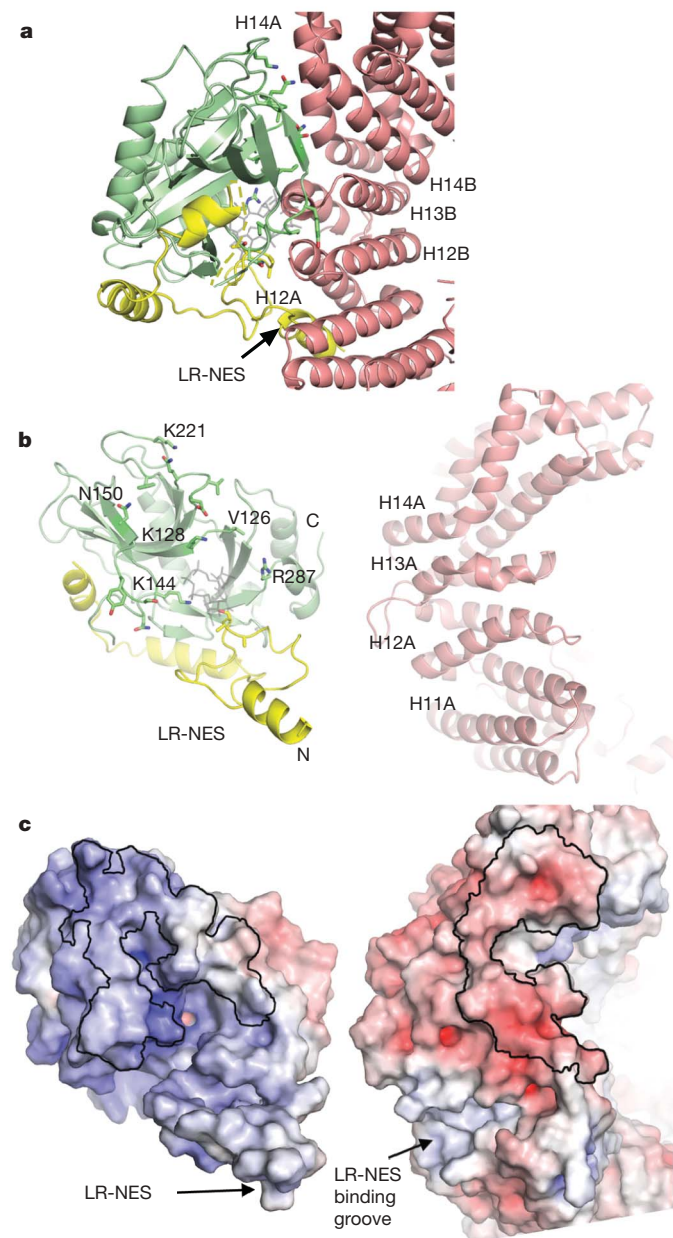


Figure 5 | Interactions of CRM1 with the SNUPN NES epitope II. **a**, Ribbon drawing of the interface between CRM1 (pink) and the NES epitope II of SNUPN, which is found on its NBD (green). The N-terminal sIBB domain of SNUPN is in yellow. No nucleotide is bound in the CRM1–SNUPN complex but the m³G-cap nucleotide from superimposed nucleotide-bound NBD (PDB accession 1XK5) is drawn in grey to show the location of the nucleotide-binding site. **b**, **c**, SNUPN and CRM1 in **a** are pried apart to show the interface (outlined with black lines) and electrostatic surface potential (scale of -12 kT e^{-1} to $+12 \text{ kT e}^{-1}$) of individual proteins at the interface.

Interactions at the second NES site are mostly polar with a basic SNUPN surface complementing a CRM1 acidic interface (Fig. 5c). Electrostatic interactions here include a salt bridge between SNUPN Lys 144 and CRM1 Asp 624, and long-range electrostatic contacts between Lys 128, Arg 129, Lys 221 and Lys 223 of SNUPN, and Asp 624, Glu 683 and Asp 681 of CRM1. The acidic CRM1 interface is only a fraction of a large acidic patch that spans the entire convex surface of repeats H13–H20 (Supplementary Fig. 10). Given the bipartite nature of the CRM1–SNUPN interaction and proximity of a large acidic surface to the LR-NES binding site, we speculate that many CRM1 substrates contain basic NES epitopes in addition to LR-NES. These other NES epitopes may be conformational epitopes

with a basic surface like the SNUPN NBD, or may be simpler linear epitopes that are enriched with basic residues.

The HIV-1 Rev protein may be another example of a multipartite CRM1 substrate. Full-length Rev binds CRM1 100-fold more tightly than its LR-NES because of extra interactions with an adjacent basic element that contacts outer helices of H15 and H17 (refs 28,38). These results indicate that at least one additional NES epitope of Rev interacts with the C-terminal acidic surface located next to the LR-NES binding groove of CRM1 (Supplementary Fig. 9). Multipartite recognition by CRM1 would increase substrate affinity beyond that of isolated LR-NESs, which generally bind weakly²¹. We have previously shown that different NLS epitopes in the multipartite PY-NLS can contribute differently to total binding energy in different substrates, and combinatorial mixing of energetically weak and strong epitopes provide a mechanism to amplify signal diversity⁴². Similarly, combination of the LR-NES with further binding epitope(s) will also allow energetic variability of individual epitopes to generate NESs that are more diverse, evolvable and tunable. Thus, similar principles of substrate recognition seem to be common for both directions of nuclear trafficking.

METHODS SUMMARY

GST–CRM1 (human, residues 1–1071) and GST–SNUPN (human, residues 1–342) were expressed separately in *Escherichia coli* BL21(DE3) and purified separately by affinity and gel filtration chromatography. The purified proteins were mixed and the CRM1–SNUPN complex was purified by gel filtration chromatography. Crystals of the CRM1–SNUPN complex were obtained by vapour diffusion using 8% PEG 8000, 0.2 M magnesium chloride, 0.1 M MES, pH 6.5, in the reservoir solution. The 2.9 Å native data ($\lambda = 0.97934 \text{ Å}$) and the MAD data from a native CRM1-selenomethionine SNUPN crystal were collected at beamline 19ID, Argonne National Laboratory and MAD phases obtained using Solve/Resolve^{43,44}. The final model has R_{free} 26.9%, R_{factor} 23.9% and contains CRM1 residues 53–57, 62–91, 96–115, 121–179, 187–387, 402–429 and 447–1061, and SNUPN residues 1–66 and 94–293.

For *in vitro* pull down assays, $\sim 15 \mu\text{g}$ GST–SNUPN or GST–NESs proteins were immobilized on $\sim 90 \mu\text{l}$ glutathione sepharose (GE Healthcare) and incubated with $\sim 2.6 \mu\text{M}$ of CRM1 proteins in the presence or absence of excess RanGTP. After extensive washing, bound proteins were separated by SDS-PAGE and Coomassie stained. To determine cellular localization of SNUPN proteins, *S. cerevisiae* BY4741 cells expressing NLS-GFP–GFP and NLS-GFP–GFP–SNUPN pRS415 plasmids were analysed as previously described⁴². pRS416 NLS-GFP–GFP–SNUPN(1–15) was transformed into the *xpo1-1* strain (plasmid KWT121; ref. 12) and BY4741 cells for the temperature-sensitive cell assay.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 December 2008; accepted 6 March 2009.

Published online 1 April 2009.

1. Tran, E. J., Bolger, T. A. & Wentz, S. R. SnapShot: nuclear transport. *Cell* **131**, 420 (2007).
2. Weis, K. Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* **112**, 441–451 (2003).
3. Gorlich, D. & Kutay, U. Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 607–660 (1999).
4. Conti, E. & Izaurralde, E. Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.* **13**, 310–319 (2001).
5. Dingwall, C., Sharnick, S. V. & Laskey, R. A. A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. *Cell* **30**, 449–458 (1982).
6. Lanford, R. E. & Butel, J. S. Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell* **37**, 801–813 (1984).
7. Kalderon, D., Richardson, W. D., Markham, A. F. & Smith, A. E. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**, 33–38 (1984).
8. Lee, B. J. *et al.* Rules for nuclear localization sequence recognition by karyopherin β 2. *Cell* **126**, 543–558 (2006).
9. Wen, W., Meinkoth, J. L., Tsien, R. Y. & Taylor, S. S. Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82**, 463–473 (1995).
10. Fischer, U. *et al.* The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**, 475–483 (1995).
11. Richards, S. A., Carey, K. L. & Macara, I. G. Requirement of guanosine triphosphate-bound ran for signal-mediated nuclear protein export. *Science* **276**, 1842–1844 (1997).

12. Stade, K., Ford, C. S., Guthrie, C. & Weis, K. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041–1050 (1997).
13. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, J. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
14. Ossareh-Nazari, B., Bachelier, F. & Dargemont, C. Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science* **278**, 141–144 (1997).
15. Fukuda, M. *et al.* CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308–311 (1997).
16. Neville, M. *et al.* The importin- β family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol.* **7**, 767–775 (1997).
17. Bogerd, H. P. *et al.* Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel *in vivo* randomization-selection assay. *Mol. Cell. Biol.* **16**, 4207–4214 (1996).
18. Henderson, B. R. & Eleftheriou, A. A comparison of the activity, sequence specificity, and CRM1-dependence of different nuclear export signals. *Exp. Cell Res.* **256**, 213–224 (2000).
19. la Cour, T. *et al.* Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.* **17**, 527–536 (2004).
20. Kutay, U. & Guttinger, S. Leucine-rich nuclear-export signals: born to be weak. *Trends Cell Biol.* **15**, 121–124 (2005).
21. Engelsma, D., Bernad, R., Calafat, J. & Fornerod, M. Supraphysiological nuclear export signals bind CRM1 independently of RanGTP and arrest at Nup358. *EMBO J.* **23**, 3643–3652 (2004).
22. Cansizoglu, A. E. *et al.* Structure-based design of a pathway-specific nuclear import inhibitor. *Nature Struct. Mol. Biol.* **14**, 452–454 (2007).
23. Mancias, J. D. & Goldberg, J. Exiting the endoplasmic reticulum. *Traffic* **6**, 278–285 (2005).
24. Swanton, E. & High, S. ER targeting signals: more than meets the eye? *Cell* **127**, 877–879 (2006).
25. Kudo, N. *et al.* Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc. Natl Acad. Sci. USA* **96**, 9112–9117 (1999).
26. Matsuyama, A. *et al.* ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*. *Nature Biotechnol.* **24**, 841–847 (2006).
27. Nishi, K. *et al.* Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. *J. Biol. Chem.* **269**, 6320–6324 (1994).
28. Paraskeva, E. *et al.* CRM1-mediated recycling of snurportin 1 to the cytoplasm. *J. Cell Biol.* **145**, 255–264 (1999).
29. Huber, J. *et al.* Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.* **17**, 4114–4126 (1998).
30. Petosa, C. *et al.* Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Mol. Cell* **16**, 761–775 (2004).
31. Strasser, A., Dickmanns, A., Luhrmann, R. & Ficner, R. Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J.* **24**, 2235–2243 (2005).
32. Mitrousis, G., Olia, A. S., Walker-Kopp, N. & Cingolani, G. Molecular basis for the recognition of snurportin 1 by importin beta. *J. Biol. Chem.* **283**, 7877–7884 (2008).
33. Dong, X., Biswas, A. & Chook, Y. M. Structural basis of assembly and disassembly of the CRM1 nuclear export complex. *Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1585 (in the press).
34. Suel, K. E., Cansizoglu, A. E. & Chook, Y. M. Atomic resolution structures in nuclear transport. *Methods* **39**, 342–355 (2006).
35. Matsuura, Y. & Stewart, M. Structural basis for the assembly of a nuclear export complex. *Nature* **432**, 872–877 (2004).
36. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
37. Meiler, J. & Baker, D. Coupled prediction of protein secondary and tertiary structure. *Proc. Natl Acad. Sci. USA* **100**, 12105–12110 (2003).
38. Askjaer, P. *et al.* The specificity of the CRM1-Rev nuclear export signal interaction is mediated by RanGTP. *J. Biol. Chem.* **273**, 33414–33422 (1998).
39. Thomas, F. & Kutay, U. Biogenesis and nuclear export of ribosomal subunits in higher eukaryotes depend on the CRM1 export pathway. *J. Cell Sci.* **116**, 2409–2419 (2003).
40. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Nuclear export signal consensus sequences defined using a localization-based yeast selection system. *Traffic* **9**, 2053–2062 (2008).
41. Englmeier, L. *et al.* RanBP3 influences interactions between CRM1 and its nuclear protein export substrates. *EMBO Rep.* **2**, 926–932 (2001).
42. Suel, K. E., Gu, H. & Chook, Y. M. Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals. *PLoS Biol.* **6**, e137 (2008).
43. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
44. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Sheng, C. Kong, D. Tomchick and C. Dann for assistance during data collection and structure determination, S. Padrick for advice on fluorescence binding assays, G. Süel and T. Cagatay for help with microscopy, K. Weis for yeast strains and constructs, X. Zhang, L. Rice and M. Rosen for discussion. This work is funded by National Institute of Health (NIH) grants R01GM069909, R01GM069909-03S1, 5-T32-GM008297, Welch Foundation grant I-1532 and the UT Southwestern Endowed Scholars Program. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source, was supported by the US Department of Energy, Office of Energy Research, under contract no. W-31-109-ENG-38.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes rcsb051649 and 3GB8. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.M.C. (yuhmin.chook@utsouthwestern.edu).

METHODS

CRM1–SNUPN complex. GST–CRM1 (human, residues 1–1071) and GST–SNUPN (human, residues 1–342) were expressed separately in *E. coli* BL21(DE3) (Stratagene) at 25 °C and purified separately. The cells were lysed in buffer A (50 mM Tris, pH 7.5, 200 mM sodium chloride, 1 mM EDTA, 2 mM DTT, 20% glycerol) followed by affinity purification with glutathione sepharose (GE Healthcare), cleavage off the column with TEV protease and gel filtration chromatography. CRM1–SNUPN mixed at 2:3 molar ratio was purified by gel filtration chromatography. Selenomethionine SNUPN was expressed in BL21(DE3) cells using M9 minimal media, purified as for native SNUPN and the native CRM1–selenomethionine SNUPN complex was purified as for the native complex.

X-ray crystallography. Crystals of the CRM1–SNUPN complex ($P_{6,22}$, $a = b = 250.4$ Å, $c = 190.1$ Å, one complex per asymmetric unit; 76% solvent) were obtained by vapour diffusion in hanging drops using 8% PEG 8000, 0.2 M magnesium chloride, 0.1 M MES, pH 6.5, in the reservoir solution at 18 °C and flash-frozen in liquid propane. Selenomethionine crystals were of the same space group and similar unit cell parameters. The 2.9 Å native data ($\lambda = 0.97934$ Å) and the MAD data ($\lambda_{\text{peak}} = 0.97935$ Å, $\lambda_{\text{inflection}} = 0.97945$ Å, $\lambda_{\text{remote}} = 0.97167$ Å) were collected at 100K at beamline 19ID, Argonne National Laboratory and processed with HKL2000 (ref. 45). Seven out of ten possible selenium sites were found using the programs Shelx⁴⁶ and Solve⁴⁴. Heavy atom refinement and MAD phasing were carried out using Solve/Resolve^{43,44}. Iterative cycles of model building using program COOT⁴⁷ and refinement using CNS⁴⁸ and a final TLS refinement⁴⁹, all using the same test data set, produced a final model ($R_{\text{free}} 26.9\%$, $R_{\text{work}} 23.9\%$) that contains CRM1 residues 53–57, 62–91, 96–115, 121–179, 187–387, 402–429 and 447–1061 and SNUPN residues 1–66 and 94–293. Average B-factor values for the model are 100.3 Å² for repeats H1 and H2 of CRM1, 37.2 Å² for H3–H20A of CRM1, 104.1 Å² for H20B to the C terminus of CRM1 and 26.8 Å² for SNUPN. Wilson B-factors of the data (64.6 Å² for the MAD inflection data) are comparable to average B-factor of the model (39.3 Å²).

In vitro binding assays. Site-directed mutagenesis of SNUPN and CRM1 were performed using the QuickChange method (Stratagene) and all constructs were sequenced. The SNUPN(1–15) NES as well as NESs of Rev and Nmd3 were generated by ligation of annealed oligonucleotides into the pGEX-Tev vector. Approximately 15 µg of immobilized GST–SNUPN or GST–NESs proteins (glutathione sepharose; GE Healthcare) were incubated with ~2.6 µM of CRM1 proteins in the presence or absence of RanGTP (fivefold molar excess). After extensive washing, bound proteins were visualized by SDS–PAGE and Coomassie staining.

Isothermal titration calorimetry experiments. The CRM1–SNUPN experiment was performed twice, with 18 and 22 injections of 5 µl SNUPN(1–342) (290–386 µM) into 1.4 ml of CRM1 (12–14 µM) at 15 °C. The CRM1–Ran experiment was performed once with 22 injections of 5 µl RanGppNHp (290 µM) into 1.4 ml

of CRM1 (12 µM) at 15 °C. All proteins were in buffer containing 50 mM Tris, pH 7.5, 145 mM sodium chloride, 5 mM magnesium chloride and 8% glycerol. The heat of dilution obtained from injecting a ligand into buffer was subtracted before K_d and binding ratio were calculated using the ORIGIN data analysis software (Microcal, Inc.). The chi-square value for each CRM1–SNUPN experiment is relatively low (7.907×10^4 and 1.410×10^5 , respectively) indicating the quality of the fit. In addition, the error (s.d.) reported in Supplementary Fig. 2 demonstrates the reproducibility of the isothermal titration calorimetry (ITC) experiments. ITC experiments were performed at 15 °C because CRM1 was more stable at a lower temperature. The CRM1–SNUPN experiment was repeated at 23 °C the data obtained at the two temperatures are similar ($K_d = 6.5$ µM, $\Delta H = -19.1$ kcal mol⁻¹ and $T\Delta S = -12.1$ kcal mol⁻¹). The fourfold weaker binding affinity at 23 °C probably reflects decreased stability of CRM1 at higher temperature.

In vivo localization studies. To clone the NLS–GFP–GFP and NLS–GFP–GFP–SNUPN plasmids, two GFP genes were cloned into a modified pRS415 (*CEN6*, *ARS*, *LEU2* and *AP^R*) shuttle vector containing a 5' ADH1 promoter⁵⁰. The SV40 NLS (PKKKRKV) was inserted upstream of the GFP genes using the Quikchange (Stratagene) method. SNUPN gene was cloned 3' of the GFP genes and GGSGG linkers were inserted by Quikchange between all four genes. The whole NLS–GFP–GFP–SNUPN(1–15) insert was also subcloned into a pRS416 (*CEN6*, *ARS*, *URA3* and *AP^R*) shuttle vector.

For the transport competition assay, BY4741 cells expressing pRS415 plasmids were grown overnight at 30 °C in SC–Leu media and analysed as previously reported⁴². For the temperature-sensitive cell assay, pRS416 NLS–GFP–GFP–SNUPN(1–15) was transformed into the *xpo1-1* strain (plasmid KWHY121; ref. 12) and BY4741 cells. Cells were grown overnight at room temperature and analysed as above. After the fluorescence was examined at room temperature, the temperature in the incubation chamber was increased to 37 °C for an hour and reporter fluorescence was examined in the same cells.

45. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
46. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
48. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
49. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
50. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).

LETTERS

Discrete sources as the origin of the Galactic X-ray ridge emission

M. Revnivtsev^{1,2}, S. Sazonov^{2,3}, E. Churazov^{2,3}, W. Forman⁴, A. Vikhlinin^{2,4} & R. Sunyaev^{2,3}

An unresolved X-ray glow (at energies above a few kiloelectronvolts) was discovered about 25 years ago and found to be coincident with the Galactic disk—the Galactic ridge X-ray emission^{1,2}. This emission^{3–10} has a spectrum characteristic of a $\sim 10^8$ K optically thin thermal plasma, with a prominent iron emission line at 6.7 keV. The gravitational well of the Galactic disk, however, is far too shallow to confine such a hot interstellar medium; instead, it would flow away at a velocity of a few thousand kilometres per second, exceeding the speed of sound in the gas. To replenish the energy losses requires a source of 10^{43} erg s⁻¹, exceeding by orders of magnitude all plausible energy sources in the Milky Way¹¹. An alternative is that the hot plasma is bound to a multitude of faint sources¹², which is supported by the recently observed similarities in the X-ray and near-infrared surface brightness distributions^{13,14} (the latter traces the Galactic stellar distribution). Here we report that at energies of ~ 6 –7 keV, more than 80 per cent of the seemingly diffuse X-ray emission is resolved into discrete sources, probably accreting white dwarfs and coronally active stars.

Observations clearly show that some fraction of the X-ray emission of the Galaxy is produced by hot, diffuse interstellar plasma, heated by, for example, supernovae¹⁵, while the bulk of previously unresolved X-ray emission at energies above 1–2 keV remains unexplained. The strong similarity of the Galactic ridge X-ray emission (GRXE) large-scale distribution and that of the near-infrared map of the Milky Way suggested a stellar origin of this emission. The stellar origin was further supported by the close agreement between the X-ray emissivity per unit stellar mass inferred for the GRXE and the collective X-ray emissivity of the stellar population within a few hundred parsecs of the Sun¹⁶.

These findings motivated us in 2008 to perform a decisive test with an ultra-deep 1 Ms observation of a small ($\sim 16 \times 16$ arcmin) field near the Galactic Centre ($l^{\text{II}} = 0.08$, $b^{\text{II}} = -1.42$) with the Chandra X-ray Observatory. We selected this region of the Galactic plane because, here, a high GRXE intensity (essential for minimizing the contribution from extragalactic sources) combines with weak interstellar absorption (crucial for maximizing the 0.5–7 keV Chandra sensitivity for discrete sources). From what we know about the Solar neighbourhood, we can expect the sources producing the bulk of the GRXE to be as faint as $\sim 10^{30}$ erg s⁻¹ and to have a surface density of 10^5 per square degree or even higher in the Galactic plane. Only with the combination of an ultra-deep exposure and the excellent angular resolution of Chandra¹⁷ (~ 0.5 arcsec) has the task of resolving the GRXE become possible.

To place the most stringent limits on the fraction of the GRXE resolvable into discrete sources, we selected a field where the telescope's angular resolution is best and spatial variations of the soft X-ray emission below 1.5 keV (which might be caused by supernova remnants) are minimal. We therefore restrict our present study to a

small circle of radius 2.56 arcmin near the telescope optical axis (see Fig. 1). We refer to this field ($l^{\text{II}} = 0.113$, $b^{\text{II}} = -1.424$) as HRES (high resolution).

The total measured X-ray surface brightness in HRES is $I_{3-7\text{ keV}} = (4.6 \pm 0.4) \times 10^{-11}$ erg s⁻¹ cm⁻² deg⁻² in the 3–7 keV band, or equivalently $I_{2-10\text{ keV}} = (8.6 \pm 0.5) \times 10^{-11}$ erg s⁻¹ cm⁻² deg⁻² in the more conventional 2–10 keV band, or, in scale-free units, $I_{2-10\text{ keV}} = 3.8 \pm 0.2$ mCrab deg⁻² (here and below the uncertainties are 68% confidence intervals that include both statistical and count rate to flux conversion uncertainties). The brightest source detected in our region has a 2–10 keV flux of $\sim 1.8 \times 10^{-14}$ erg s⁻¹ cm⁻² and thus a luminosity of $\sim 10^{32}$ erg s⁻¹ if it is located at approximately the Galactic Centre distance (~ 8 kpc). More luminous, rarer sources are found in our Chandra field, but outside HRES; we exclude such sources ($I_{2-10\text{ keV}} > 10^{32}$ erg s⁻¹) from consideration when addressing the resolved fraction of the GRXE below.

The total measured X-ray surface brightness must include the contribution from the nearly isotropic extragalactic cosmic X-ray background (CXB¹⁸). The mean CXB intensity over the sky measured by Chandra in the 2–10 keV energy band¹⁹ is $I_{\text{CXB}, 2-10\text{ keV}} = 2.19 \times 10^{-11}$ erg s⁻¹ cm⁻² deg⁻², of which 31% is provided²⁰ by sources (mostly active galactic nuclei and quasars) brighter than 2×10^{-14} erg s⁻¹ cm⁻². Given the absence of such bright sources in HRES, the total CXB contribution is $\sim 1.5 \times 10^{-11}$ erg s⁻¹ cm⁻² deg⁻². After subtraction of this extragalactic emission, the GRXE intensity in HRES is $I_{\text{GRXE}, 2-10\text{ keV}} = (7.1 \pm 0.5) \times 10^{-11}$ erg s⁻¹ cm⁻² deg⁻².

Looking at the same field in the near-infrared band, which provides the best window on the Galactic stellar mass distribution, the 3.5 μ m intensity measured with the Spitzer IRAC instrument is 21 ± 2 MJy sr⁻¹ (the uncertainty being mainly due to the variance of the number of bright near-infrared point sources within the small area of the study). Given the interstellar extinction towards HRES $A_V \approx 3.5$ –4.5 (ref. 21) and adopting $A_{3.5\mu\text{m}}/A_V = 0.066$ (refs 21, 22), the extinction-corrected near-infrared surface brightness is $I_{3.5\mu\text{m}} = 26$ –29 MJy sr⁻¹. Therefore, the GRXE to near-infrared intensity ratio in HRES is $I_{2-10\text{ keV}}$ (in units of 10^{-11} erg s⁻¹ cm⁻² deg⁻²)/ $I_{3.5\mu\text{m}}$ (in units of MJy sr⁻¹) = 0.25 ± 0.04 , in excellent agreement with the value characterizing the entire Galaxy, 0.26 ± 0.05 , deduced from large-scale mapping of the GRXE¹³. This confirms that the findings of the present study of a tiny region of the Galaxy may be regarded as representative of the GRXE as a whole.

We have detected sources in the broad 0.5–7 keV energy band in the summed image of the HRES region (see Fig. 1). The sensitivity limit $f_{\text{lim}} \approx 10^{-16}$ erg s⁻¹ cm⁻² (minimum detectable flux in the 0.5–7 keV band corrected for the interstellar absorption) corresponds to a minimum detectable luminosity of $L_{0.5-7\text{ keV}} \approx 10^{30}$ erg s⁻¹ at a source distance of 8 kpc, where most of the Galactic objects in this field are expected to reside. In total, 473 sources have been detected

¹Excellence Cluster Universe, Technische Universität München, 85748, Garching, Germany. ²Space Research Institute, 117997, Moscow, Russia. ³Max-Planck-Institut für Astrophysik, 85748, Garching, Germany. ⁴Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.

with statistical significance $>4\sigma$ (minimum number of counts per source is about 10). In Fig. 2a we show the energy spectrum of the total emission from HRES, as well as the two components associated with the detected sources and with the remaining unresolved emission. Most importantly, the summed spectrum of detected sources exhibits a pronounced ~ 6.7 keV iron emission line, a distinctive feature of the GRXE which was often regarded as an important argument in favour of it being the emission of a truly diffuse hot plasma^{3,11}. But now we clearly see that the bulk of the 6.7 keV line emission, as well as of the neighbouring continuum, is in fact produced by point sources. We note that apart from the dominant 6.7 keV line, the unresolved (partially due to finite energy resolution of the instrument and due to limited statistics of the observation) blend of lines at 6–7 keV may contain some contribution from 6.4 keV iron fluorescent emission, part of which may be unrelated to the GRXE and result from irradiation of the interstellar medium by discrete X-ray sources^{23,24}.

The derived fraction of the X-ray emission resolved into point sources is shown as a function of energy in Fig. 2b. In the narrow

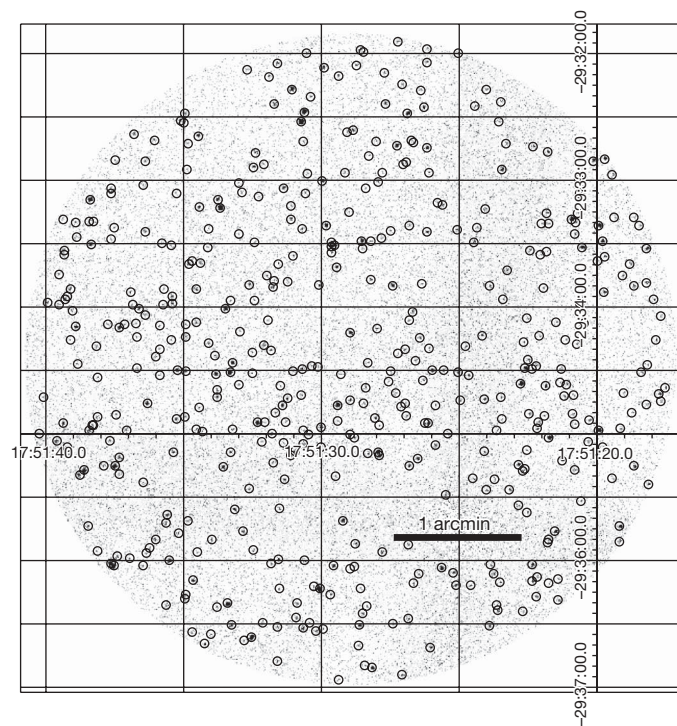


Figure 1 | The Chandra image in the 0.5–7 keV energy band. Circles of radius 2 arcsec denote the positions of point sources detected after 1 Ms exposure. The Chandra data were reduced following a standard procedure²⁵. The detector background was modelled using the stowed data set (<http://cxc.harvard.edu/contrib/maxim/stowed>) and adjusted to the conditions of the current observations using the count rate at energies 9–12 keV, where Chandra has almost zero effective area. The total measured X-ray surface brightness in HRES is $I_{3-7 \text{ keV}} = (4.6 \pm 0.4) \times 10^{-11} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ deg}^{-2}$ in the 3–7 keV band, or equivalently $I_{2-10 \text{ keV}} = (8.6 \pm 0.5) \times 10^{-11} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ deg}^{-2}$ in the 2–10 keV band. Throughout the field there are noticeable variations of the soft X-ray (<2 keV) surface brightness, due to what appears to be a previously unknown supernova remnant shell projected onto the Chandra field. We note that if a 1 Ms Chandra observation were repeated in a nearby field, the measured X-ray surface brightness would be slightly different because the number of brightest point sources varies from field to field, an effect known as cosmic variance in extragalactic studies. For the same reason, there may be subtle field-to-field variations in the GRXE spectral shape, and in particular in emission line ratios, and recent observations indicated that such variations do exist¹⁰. Additional variations of the spectrum of the unresolved Galactic X-ray emission can be caused by the presence of genuine diffuse X-ray emitters such as supernova remnants.

energy band 6.5–7.1 keV that contains the iron emission line, $84 \pm 12\%$ of the total X-ray emission is resolved. Moreover, we recall that the remaining unresolved X-ray emission contains a non-negligible contribution from the CXB. Assuming that the intensity of this unresolved component in our 1 Ms Chandra observation is the same as in the Chandra extragalactic deep fields¹⁹ ($I_{\text{CXB, unres 1 Ms}} = (3.4 \pm 1.7) \times 10^{-12} \text{ erg s cm}^{-2} \text{ deg}^{-2}$ in the 2–8 keV energy band, or $(2.9 \pm 1.4) \times 10^{-13} \text{ erg s cm}^{-2} \text{ deg}^{-2}$ at 6.5–7.1 keV, assuming a power-law spectral shape with $\Gamma = 1.4$), we can estimate that $4 \pm 2\%$ of the total intensity in the 6.5–7.1 keV band is unresolved CXB emission. We conclude that we have resolved as much as $88 \pm 12\%$ of the GRXE emission into point sources at energies near the 6.7 keV line, the feature that was previously used as the strongest argument in favour of a diffuse origin for the GRXE.

Apart from a small contribution from extragalactic sources (about 40–50 sources out of 473), most of the sources detected by Chandra in HRES are probably accreting white dwarfs (with luminosities $L_{2-10 \text{ keV}} \approx 10^{31} - 10^{32} \text{ erg s}^{-1}$) and binary stars with strong coronal activity (with $L_{2-10 \text{ keV}} < 10^{31} \text{ erg s}^{-1}$). Indeed, if we plot the fraction of the total GRXE flux contained in sources with fluxes higher than a variable detection threshold (see Fig. 3), the resulting dependence proves to be in good agreement with the expectation based on the luminosity function of faint X-ray sources measured in the Solar

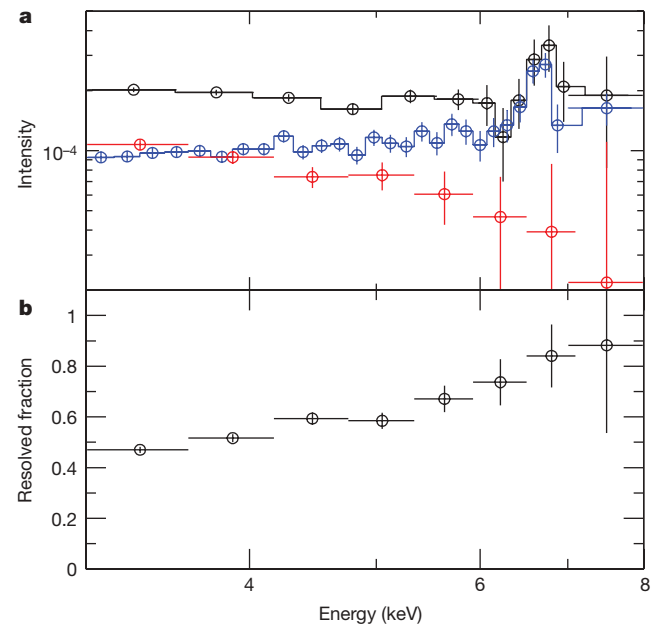


Figure 2 | GRXE spectrum and its resolved fraction. **a**, Spectra collected by Chandra within the HRES region. Black data points, error bars and the histogram show the spectrum of the total emission from HRES; the collective spectrum of all detected sources is presented in blue and the spectrum of the remaining unresolved emission in the current observations is in red. The integrated spectrum of detected sources exhibits a strong ~ 6.7 keV iron emission line, characteristic of hot (with temperatures $10 - 100 \times 10^6$ K) plasma emission. This line has been the main support for the popular hypothesis that the GRXE has a truly diffuse, interstellar origin, even though such hot interstellar plasma cannot be confined within the Galaxy by its gravitational potential. We took into account that a small fraction of photons, X (10% at energies 4–6 keV, according to the Chandra Proposers' Observatory Guide²⁶) from a point source are scattered by the telescope outside the surrounding circle of radius 2 arcsec. We therefore corrected the directly measured collective spectrum of detected sources $F_1(E)$ using the formula $\bar{F}_1(E) = [F_1(E) - F_2(E)A_1/A_2]/[1 - X - XA_1/A_2]$, where $F_2(E)$ is the spectrum of the unresolved X-ray emission, A_1 ($\sim 2\%$ of the total) is the area covered by the circles of radius 2 arcsec used for collecting the source fluxes, and A_2 is the area outside these circles. **b**, Fraction of the X-ray emission resolved by Chandra into point sources as a function of X-ray photon energy. Error bars in **a** and **b** are 68% confidence intervals.

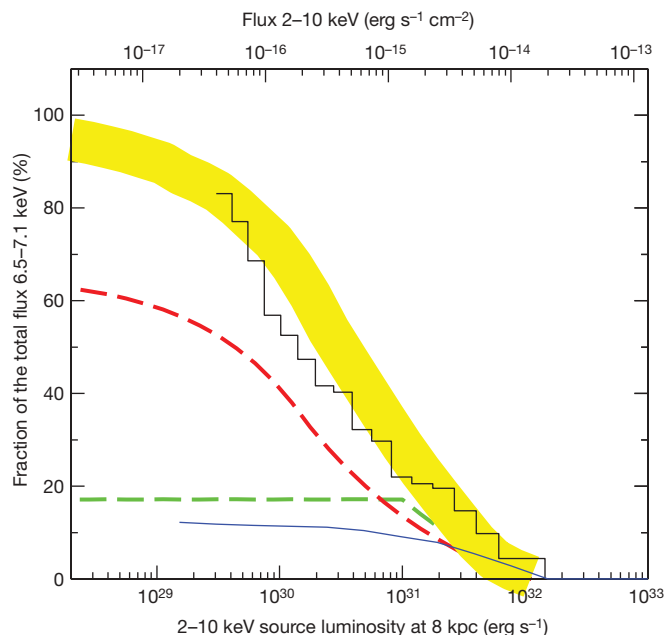


Figure 3 | Fraction of resolved X-ray emission around the 6.7 keV iron emission line as a function of the limiting source flux/luminosity. The histogram shows the fraction of the total flux in the 6.5–7.1 keV energy band in the Chandra field provided by discrete sources with fluxes above a given detection threshold in the 2–10 keV energy band. The wide yellow curve shows the corresponding dependence expected for a combination of Galactic sources with luminosities below $10^{32} \text{ erg s}^{-1}$ (all located at the Galactic Centre distance of 8 kpc) and extragalactic sources with fluxes below $2 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ in this energy band. The blue curve shows the expected contribution of extragalactic sources (mostly active galactic nuclei)²⁰. The green and red curves show the expected contributions of accreting white dwarfs and coronally active stars correspondingly (X-ray spectra of these types of sources are described in refs 27 and 28, for example), estimated using the luminosity functions of these classes of objects measured in the Solar vicinity¹⁶. These Galactic curves were normalized so that the total resolved fraction given by the model is equal to the actual measured fraction of 84% at the detection limit of the 1 Ms Chandra observation of $10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$. Fluxes of sources detected in the total Chandra band 0.5–7 keV were converted into the 2–10 keV energy band using a count-rate-dependent conversion factor that was estimated using stacked Chandra spectra of bright, medium and faint sources (>100 , 10–100 and 5–10 net counts in the image, respectively). For Galactic sources, their fluxes in the 2–10 keV band were converted to the 6.5–7.1 keV band using a single conversion factor estimated from the stacked Chandra spectrum of all detected sources. Comparison of the stacked spectra of bright, medium and faint sources indicates that this conversion factor does not vary by more than a factor of 1.3 around the adopted value. Conversion of fluxes of extragalactic sources from 2–10 keV to 6.5–7.1 keV was done assuming that they have power-law spectra with photon index $\Gamma = 1.4$.

vicinity¹⁶. Furthermore, because this locally determined luminosity function continues to rise towards lower luminosities, we can expect the still unresolved ~ 10 –20% of the GRXE flux also to be composed of coronally active and normal (Sun-like) stars with luminosities $L_{2-10 \text{ keV}} < 4 \times 10^{29} \text{ erg s}^{-1}$, which are too weak to be detected at the Galactic Centre distance even in ultra-deep Chandra exposures. The contribution of such faint stellar sources should rise with the decrease of the photon energies because they typically have quite soft spectra. This might be one of the reasons why we resolve more flux at high energies than at lower energies.

The final resolution of the GRXE into discrete sources has far-reaching consequences for our understanding of a variety of astrophysical phenomena. Apart from the removal of a major energy puzzle for the Galaxy, the important immediate outcome is that we

can now use the GRXE as a measure of the cumulative emission of faint Galactic X-ray sources in the sense that spatial variations of the GRXE properties over the Milky Way can indicate intrinsic variations in the stellar populations. It has also become clear that the apparently diffuse X-ray emission of external galaxies must contain, and in some cases be dominated by, unresolved emission from faint stellar-type sources, namely accreting white dwarfs and coronally active stars.

Received 4 December 2008; accepted 5 March 2009.

- Worrall, D. M. *et al.* HEAO 1 measurements of the Galactic ridge. *Astrophys. J.* **255**, 111–121 (1982).
- Warwick, R. S. *et al.* The Galactic ridge observed by EXOSAT. *Nature* **317**, 218–221 (1985).
- Koyama, K. *et al.* Thermal X-ray emission with intense 6.7-keV iron line from the Galactic ridge. *Publ. Astron. Soc. Jpn* **38**, 121–131 (1986).
- Yamauchi, S. & Koyama, K. The 6.7 keV iron line distribution in the Galaxy. *Astrophys. J.* **404**, 620–624 (1993).
- Yamauchi, S. *et al.* Unresolved X-ray emission from the Galactic ridge with ASCA. *Publ. Astron. Soc. Jpn* **48**, L15–L20 (1996).
- Hands, A. D. P. *et al.* X-ray source populations in the Galactic plane. *Mon. Not. R. Astron. Soc.* **351**, 31–56 (2004).
- Ebisawa, K. *et al.* Chandra deep X-ray observation of a typical Galactic plane region and near-infrared identification. *Astrophys. J.* **635**, 214–242 (2005).
- Koyama, K. *et al.* Iron and nickel line diagnostics for the Galactic Center diffuse emission. *Publ. Astron. Soc. Jpn* **59**, 245–255 (2007).
- Ebisawa, K. *et al.* Spectral study of the Galactic ridge X-ray emission with Suzaku. *Publ. Astron. Soc. Jpn* **60**, 223–230 (2008).
- Yamauchi, S. *et al.* Iron emission lines on the Galactic ridge observed with Suzaku. Preprint at <<http://arxiv.org/abs/0810.0317>> (2008).
- Tanaka, Y. ASCA observation of X-ray emission from the Galactic ridge. *Astron. Astrophys.* **382**, 1052–1060 (2002).
- Worrall, D. M. & Marshall, F. E. Stellar contributions to the hard X-ray Galactic ridge. *Astrophys. J.* **267**, 691–697 (1983).
- Revnivtsev, M. *et al.* Origin of the Galactic ridge X-ray emission. *Astron. Astrophys.* **452**, 169–178 (2006).
- Revnivtsev, M., Molokov, S. & Sazonov, S. Map of the Galaxy in the 6.7-keV emission line. *Mon. Not. R. Astron. Soc.* **373**, L11–L15 (2006).
- Park, S. *et al.* Characteristics of diffuse X-ray line emission within 20 parsecs of the Galactic Center. *Astrophys. J.* **603**, 548–559 (2004).
- Sazonov, S. *et al.* X-ray luminosity function of faint point sources in the Milky Way. *Astron. Astrophys.* **450**, 117–128 (2006).
- Weisskopf, M. C. *et al.* An overview of the performance and scientific results from the Chandra X-Ray Observatory. *Publ. Astron. Soc. Jpn* **114**, 1–24 (2002).
- Giacconi, R. *et al.* Evidence for X-rays from sources outside the Solar System. *Phys. Rev. Lett.* **9**, 439–443 (1962).
- Hickox, R. C. & Markevitch, M. Absolute measurement of the unresolved cosmic X-ray background in the 0.5–8 keV band with Chandra. *Astrophys. J.* **645**, 95–114 (2006).
- Moretti, A. *et al.* The resolved fraction of the cosmic X-ray background. *Astrophys. J.* **588**, 696–703 (2003).
- Dutra, C. M. *et al.* Extinction within 10° of the Galactic Centre using 2MASS. *Mon. Not. R. Astron. Soc.* **338**, 253–262 (2003).
- Indebetouw, R. *et al.* The wavelength dependence of interstellar extinction from 1.25 to 8.0 μm using GLIMPSE data. *Astrophys. J.* **619**, 931–938 (2005).
- Sunyaev, R. A., Markevitch, M. & Pavlinsky, M. The center of the Galaxy in the recent past—a view from GRANAT. *Astrophys. J.* **407**, 606–610 (1993).
- Koyama, K. *et al.* ASCA view of our Galactic Center: remains of past activities in X-rays? *Publ. Astron. Soc. Jpn* **48**, 249–255 (1996).
- Vikhlinin, A. *et al.* Chandra temperature profiles for a sample of nearby relaxed galaxy clusters. *Astrophys. J.* **628**, 655–672 (2005).
- Chandra. Proposers' Observatory Guide <<http://cxc.harvard.edu/proposer/POG/html/index.html>> (23 January 2009).
- Hellier, C., Mukai, K. & Osborne, J. P. Iron K α linewidths in magnetic cataclysmic variables. *Mon. Not. R. Astron. Soc.* **297**, 526–530 (1998).
- Huenemoerder, D., Canizares, C. & Schulz, N. X-ray spectroscopy of II Pegasi: coronal temperature structure, abundances, and variability. *Astrophys. J.* **559**, 1135–1146 (2001).

Acknowledgements M.R. thanks M. Markevitch for his help with the Chandra instrumental background. This research was supported by the DFG Cluster of Excellence “Origin and Structure of the Universe”, by NASA Chandra grant GO8-9132A, by the OFH-17 programme of the Russian Academy of Sciences, and by grants RFFI 07-02-01004 and RFFI 07-02-00961.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.R. (mikej@mpa-garching.mpg.de).

Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena

K. Goda^{1*}, K. K. Tsia^{1*} & B. Jalali^{1*}

Ultrafast real-time optical imaging is an indispensable tool for studying dynamical events such as shock waves^{1,2}, chemical dynamics in living cells^{3,4}, neural activity^{5,6}, laser surgery^{7–9} and microfluidics^{10,11}. However, conventional CCDs (charge-coupled devices) and their complementary metal–oxide–semiconductor (CMOS) counterparts are incapable of capturing fast dynamical processes with high sensitivity and resolution. This is due in part to a technological limitation—it takes time to read out the data from sensor arrays. Also, there is the fundamental compromise between sensitivity and frame rate; at high frame rates, fewer photons are collected during each frame—a problem that affects nearly all optical imaging systems. Here we report an imaging method that overcomes these limitations and offers frame rates that are at least 1,000 times faster than those of conventional CCDs. Our technique maps a two-dimensional (2D) image into a serial time-domain data stream and simultaneously amplifies the image in the optical domain. We capture an entire 2D image using a single-pixel photodetector and achieve a net image amplification of 25 dB (a factor of 316). This overcomes the compromise between sensitivity and frame rate without resorting to cooling and high-intensity illumination. As a proof of concept, we perform continuous real-time imaging at a frame speed of 163 ns (a frame rate of 6.1 MHz) and a shutter speed of 440 ps. We also demonstrate real-time imaging of microfluidic flow and phase-explosion effects that occur during laser ablation.

Optical imaging is a widespread and versatile diagnostics and inspection tool in use today. Although most of the current research in imaging is aimed at improving the spatial resolution to below the diffraction limit^{12,13}, there are numerous applications that demand improvement in temporal resolution. Imaging systems with high temporal resolution are needed to study rapid physical phenomena ranging from shock waves, including extracorporeal shock waves used for surgery¹, to diagnostics of laser fusion² and fuel injection in internal combustion engines. High-speed imaging is becoming increasingly important in microscopy because on the micrometre scale even slow-moving phenomena require high temporal resolution. One example is the spatiotemporal study of biochemical waves in cells and tissues, which requires imaging with a micro- to nano-second response time and is important for the study of cell signalling and drug transport¹⁴. Another key application is in the field of flow cytometry¹¹, where high-speed cameras are needed to provide high-throughput cell characterization.

The CCD or CMOS imager is by far the most widely deployed optical imaging technology. It offers a spatial resolution of a few micrometres, a large number of pixels and relatively low cost. Typical imagers used in consumer electronics have frame rates of 30 Hz, although high-end versions can operate at rates on the order of 1 kHz by reducing the number of pixels that are read out from the arrays. A CCD imager

with a 1-MHz frame rate—the world's fastest CCD camera—has been developed by Shimadzu Corporation (ref. 15; see also <http://www.shimadzu.com/products/test/hsvc/oh80jt0000001d6t-att/booe13-02.pdf>). Its impressive performance stems from several technological modifications aimed at overcoming the loss of sensitivity that results from the shorter integration time during high-speed imaging, and also at reducing the time it takes to read out the data from the 2D pixel arrays. These modifications include a larger pixel size for the purpose of adding pixel-level *in situ* memory (at the expense of reduced spatial resolution), and cooling of the camera to reduce noise (at the expense of the added complexity of refrigeration). The Shimadzu device also relies on a high-power illuminator to ensure adequate signal-to-noise ratio and to prevent a drop in sensitivity—a requirement that renders it unsuitable for microscopy, where focusing of the high-power illumination over a small field of view can cause damage to a biological sample. This compromise between sensitivity and frame rate is not unique to the CCD—it impacts almost all imaging and detection systems.

In scientific applications, high-speed imaging is often achieved using the time-resolved pump–probe technique^{16–18}. Pump–probe techniques can capture the dynamics of fast events, but only if the event is repetitive. Because they do not operate in real time, they are unable to capture non-repetitive random events that occur only once or do not occur at regular intervals, such as rogue events¹⁹. Detection of such events requires an imaging technology with fast, continuous, real-time capability.

Another type of high-speed image sensor is the framing streak camera, which has been used for diagnostics in laser fusion, plasma radiation and combustion. This device operates in burst mode only (providing only several frames; see <http://learn.hamamatsu.com/tutorials/streakcamera/>) and requires synchronization of the camera with the event to be captured, rendering streak cameras also unable to capture unique or random events. This, along with the high cost of the camera, limits its use in practical applications.

Although CCD imagers will continue to be the most widely used imaging modality and pump–probe experiments will remain a powerful tool for studying fast repetitive events, a new and complementary imaging modality that can capture the dynamics of fast single-shot or random events is clearly needed. Serial time-encoded amplified microscopy (STEAM) technology is a new approach to imaging that provides such a capability. The 2D image is encoded into a serial time-domain waveform that is amplified and captured, not by a CCD camera, but instead by a single-pixel photodiode and an oscilloscope. The main attributes of the new imager are the image amplification in the optical domain and the elimination of the CCD—when combined, they enable continuous real-time operation at a frame rate of 6.1 MHz and a shutter speed of 440 ps. The STEAM camera operates continuously and can capture ultrafast events

¹Department of Electrical Engineering, University of California, Los Angeles, California 90095, USA.

*These authors contributed equally to this work.

without any knowledge of the timing of their occurrence. The throughput of the STEAM camera can be further increased with minor modifications to the system shown in Fig. 1 (see Methods and Supplementary Information.)

The key feature of the STEAM camera is the mapping of a 2D image into an amplified serial time-domain waveform. This is achieved by first encoding the 2D spatial information of an object onto the spectrum of a broadband (continuum) pulse (Fig. 1) using a pair of orthogonally oriented spatial dispersers (Methods). A similar component has previously been used for spectroscopy²⁰. The encoding occurs when the spatially dispersed pulse is reflected off the object, after which it re-enters the disperser, where the wavelengths are recombined. An optical circulator directs the pulses into an optically amplified dispersive Fourier transformer, the details of which are shown in the inset of Fig. 1.

Image amplification is performed in the optical domain and occurs in two stages. An erbium-doped fibre amplifier pre-amplifies the image by 10 dB (a factor of ten). The image then enters a dispersive fibre whose function is twofold. By virtue of high group-velocity dispersion (also called temporal dispersion), the dispersive fibre performs a Fourier transform operation, mapping the spectrum into the time domain. The optical spectrum, which is encoded with the spatial image, now appears as a serial sequence in time (Fig. 2a). To amplify the image simultaneously (second-stage amplification), the dispersive fibre is pumped with secondary light sources to implement stimulated Raman amplification directly within the dispersive medium. This technique is crucial because it not only compensates for the loss inherent in the dispersive fibre, but also provides an additional net image amplification of 15 dB (a factor of 31.6). Therefore, the system amplifies the image by a total of 25 dB (a factor of 316) in the optical domain.

Dispersion-compensating fibre (DCF) is used to perform the dispersive Fourier transform and Raman amplification of the image because it offers high dispersion and low loss as well as a high Raman gain coefficient^{21–23}. Broadband gain can be achieved by pumping with multiple lasers or an incoherent light source²². As the image enters the DCF, the aperture of the fibre rejects the scattered light from out-of-focus planes, rendering the system a confocal microscope.

The amplified dispersive Fourier transform maps the spectrum of an optical pulse into a time-domain waveform, allowing the spectrum to be viewed on a conventional electronic receiver (Fig. 2a). It

eliminates the traditional spectrometer, including the prism or diffraction grating and the CCD, and makes it possible to capture the spectrum with a single-pixel photodetector. At the same time, it overcomes the receiver noise level by amplifying the signal while it is still in the optical domain. Optical image amplification is vital for high-speed imaging because the noise power present in any measurement increases inversely with the integration time. Without this amplification, the image would not be visible because the signal would lie below the thermal noise level of the photodetector, as can be seen in Fig. 2b. A high-power laser source can also overcome the detector noise; however, this is not a desirable solution as it can potentially damage the object being imaged, particularly in microscopy, in which the light has to be focused onto a very small area. The optical image amplification made possible by the serialization of the 2D image overcomes this problem.

The image amplifier in our STEAM camera is different from ‘image intensifiers’. In STEAM, amplification occurs in the optical domain, whereas in image intensifiers, it occurs in the electronic domain. Image intensifiers are complex devices and have a low image-acquisition rate of up to ~ 10 kHz in continuous mode^{24,25}, which is adequate for their intended use in night-vision cameras because these only need to operate at the video rate. The limited frame rate is due to the fundamental compromise between gain and bandwidth in all electronic systems, including the image intensifier.

Operation of the STEAM camera is illustrated in Fig. 2c. The figure shows the image of a 2D object captured in the time domain using a real-time oscilloscope. The reconstruction of the 2D spatial image from a one-dimensional (1D) temporal data stream involves the trivial task of simply sorting the 1D vector into a 2D matrix and is described further in Methods. The results clearly demonstrate a modality for optical imaging that is amenable to high-speed operation. The 2D frame corresponds to the envelope of the temporal pulse shown in Fig. 2a. Frames are repeated at intervals of 163 ns (Fig. 2d)—a value that corresponds to a frame rate of 6.1 MHz, which is a record for continuous real-time imaging. The shutter speed is 440 ps (Methods), which is also a record for continuous real-time imaging. For comparison, Fig. 2c shows the image of the same object using a high-sensitivity infrared camera which has a frame rate of 50 Hz, about 120,000 times less than that of the STEAM camera. In this experiment, the 2D dispersed light traverses the object and the image is captured on the camera in its position behind the object.

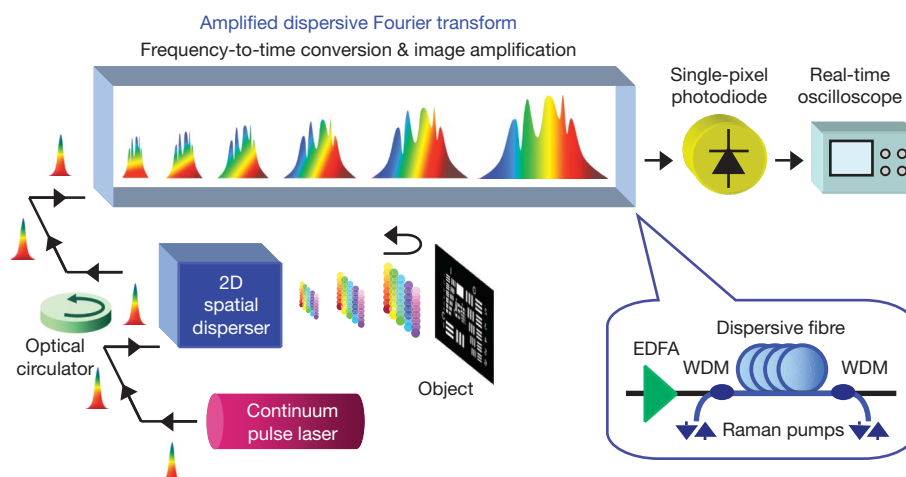


Figure 1 | STEAM camera. The STEAM camera maps a 2D spatial image into a serial time-domain waveform. It starts by encoding the 2D image into the spectrum of a broadband (continuum) pulse (Methods). The optical circulator directs the pulse reflected off the object into an optically amplified dispersive Fourier transformer (inset). An erbium-doped fibre amplifier (EDFA) pre-amplifies the image by 10 dB. The image then enters a dispersive fibre, where it is further boosted by 15 dB using a distributed Raman amplifier pumped by multiple powerful lasers through wavelength-division

multiplexers (WDMs), and its spectrum is simultaneously mapped into the time domain. The optical spectrum appears as a serial sequence in time, allowing the image to be captured with a single-pixel photodiode and an oscilloscope. This scheme eliminates the need for a CCD and also performs an image amplification of 25 dB in the optical domain. The amplification enables fast real-time imaging of dynamic events. The STEAM camera can also be used as a confocal microscope (see text). Supplementary Movie 1 illustrates the functionality of the STEAM camera.

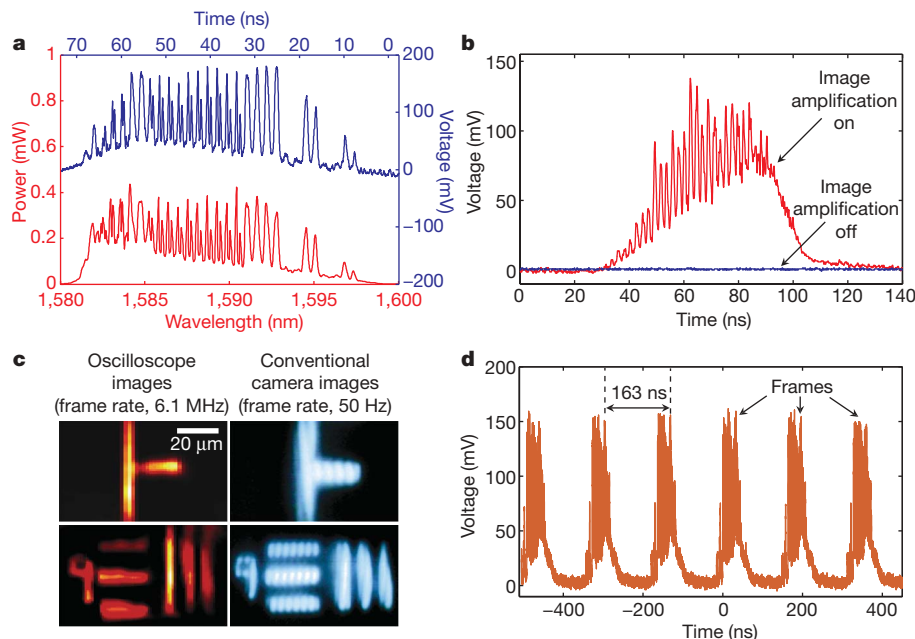


Figure 2 | Basic operation of the STEAM camera. **a**, Comparison between the temporal waveform (single pulse) captured by the oscilloscope (blue) and the spectrum measured by a conventional optical spectrum analyser (red). The consistency between them validates the frequency-to-time mapping operation performed by the dispersive Fourier transformation. **b**, The temporal waveform, representing the image, with (red) and without (blue) the image amplification. The optical amplification raises the otherwise undetectably weak signal significantly above the detector noise floor, overcoming the fundamental compromise between imaging sensitivity

and frame rate. **c**, Images of a calibration target captured by the STEAM camera, compared with images obtained using a conventional infrared camera. The shutter speed is estimated to be 440 ps. The granularity in the horizontal direction is an artefact of the particular implementation of the 2D spatial disperser, and is not a fundamental characteristic of the system. **d**, The pulse train captured by the single-pixel photodiode and displayed on the real-time digitizer (oscilloscope). The gap between consecutive pulses is 163 ns. Each pulse contains one frame of the 2D image.

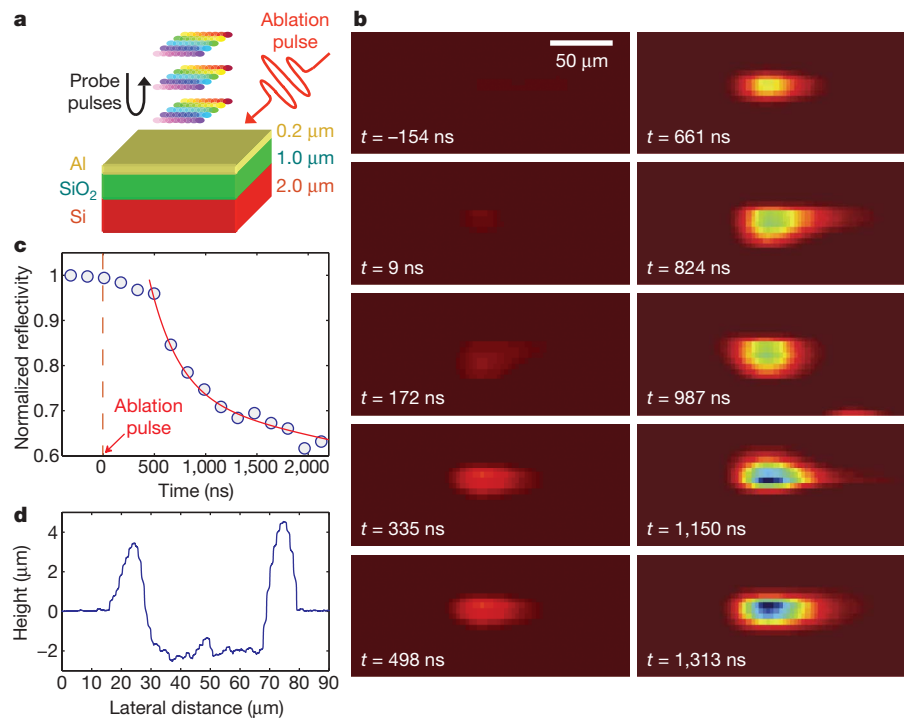


Figure 3 | Schematic of experimental set-up, real-time STEAM images and characterization of the laser ablation experiment. **a**, Schematic set-up for the laser ablation experiment. A mid-infrared pulse laser (6-mJ pulse energy, 5-ns pulse width) is focused at an angle onto the sample with a bilayer of aluminium and silicon dioxide deposited on top of a silicon-on-insulator substrate, while STEAM pulses monitor the ablation process at normal incidence. **b**, Real-time 2D images captured by the STEAM camera with a temporal resolution of 163 ns and shutter speed of 440 ps. The changes in sample surface reflectivity due to the laser-induced mass ejection are evident

after the excitation pulse hits the sample at $t = 0$ ns. **c**, The time-sequenced surface reflectivity change showing that the time delay between the pulse excitation and the sudden decrease in the surface reflectivity correlates with the mass-ejection process. This phenomenon is a clear signature of the phase-explosion effect that occurred in the sample. **d**, The depth profile of the sample, obtained with a profilometer, indicating that the ablation pulse ablated the aluminium and silicon dioxide layers and exposed the silicon layer.

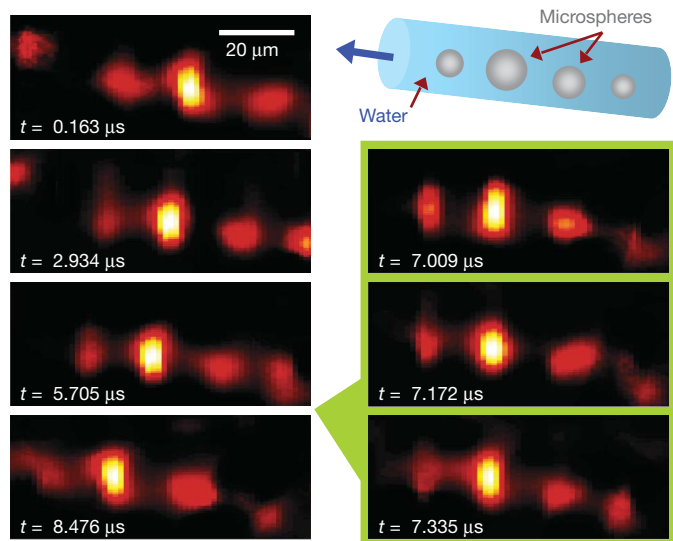


Figure 4 | Schematic and real-time STEAM images of the microfluidics experiment. An ultrafast microfluidic flow of water-suspended metal microspheres (sphere diameters ranging from 10 to 30 μm) was established in a hollow fibre with an inner diameter of 50 μm . On the left, one out of every seventeen snapshots is shown, for clarity. On the right, consecutive snapshots with the full temporal resolution of 163 ns and shutter speed of 440 ps are shown. A total of 3,000 frames can be captured, limited by the memory of the real-time digitizer used in the experiment. A flow of microspheres at a flow speed of about 2.4 m s^{-1} from right to left can be seen. Supplementary Movie 2 is a real-time movie of the microfluidic flow.

As a demonstration of the ultrafast real-time imaging capability of the STEAM camera, we used it to capture the dynamics of laser ablation. Laser ablation is a ubiquitous technology that is used in laser surgery, laser cutting and micromachining, and laser-induced breakdown spectroscopy. The ablation was performed with a mid-infrared pulse laser (with a pulse energy of 6 mJ and a pulse width of 5 ns) focused at an angle onto a sample consisting of a bilayer of aluminium and silicon dioxide deposited on top of a silicon-on-insulator substrate. The imaging pulse train of the STEAM camera was normally incident to the surface of the sample (Fig. 3a). Figure 3b shows the real-time sequence of the images with a frame-repetition period of 163 ns. The entire frame sequence corresponding to the dynamics (laser-induced mass ejection) caused by the single ablation pulse was captured in real time. Further analysis of the surface reflectivity change shows there to be a finite time delay between the pulse excitation and the sudden decrease in the surface reflectivity, which correlates with ejection of material from the sample (Fig. 3c) and is a clear signature of the phase-explosion effect that is the hallmark of laser ablation¹⁸. Our results also show the depth profile of the sample, and confirm that the laser excitation pulse ablated the aluminium and silicon dioxide layers and exposed the underlying silicon layer. This demonstration firmly establishes the ability of the STEAM camera to monitor fast dynamical processes in real time.

As further proof of the imager's capability, we used the STEAM camera to capture fast microfluidic flow. We set up fast microfluidic laminar flow of water-suspended metal microspheres, with diameters ranging from 10 to 30 μm , in a hollow fibre with a diameter of 50 μm . Figure 4 shows the images captured by the STEAM camera. A total of 3,000 frames were captured in real time, limited only by the memory of the oscilloscope. On the left of the figure, one out of every seventeen snapshots is shown, for clarity; on the right, consecutive frames with the frame resolution of 163 ns are shown. The flow of metal microspheres from the right to the left is clear, and was calculated to have a speed of 2.4 m s^{-1} . A video of the microfluidic flow is available (Supplementary Movie 2). The temporal resolution here is the highest yet achieved in the observation of such ultrafast microfluidic flow.

This capability could prove useful in the development of microfluidic biochips, which have the potential to revolutionize cytometry and analysis techniques in molecular biology.

METHODS SUMMARY

The optical source we used was a mode-locked femtosecond laser. The centre wavelength, bandwidth, pulse energy and pulse-repetition rate of the laser after spectral broadening, filtering and pulse picking were 1,590 nm, 15 nm, 82 pJ and 6.1 MHz, respectively. The 2D spectral pattern was produced by the 2D spatial disperser and focused onto the sample using an objective lens. This is an extension of previously demonstrated 1D spectrally encoded imaging^{23,26,27}, and eliminates the need for mechanical scanning in the second dimension. The 2D disperser comprised a pair of orthogonally oriented dispersers, which consisted of a diffraction grating with a groove density of 1,200 lines per millimetre and a virtually imaged phased array with a free spectral range of 67 GHz and a linewidth of 550 MHz. Such 2D dispersers have previously been used for spectroscopy²⁰ and wavelength demultiplexing in optical data communications²⁸.

The amplified dispersive Fourier transformer consisted of a DCF with a total group-velocity dispersion of -3.3 ns nm^{-1} , four $\sim 300\text{-mW}$ continuous-wave pump lasers at 1,470–1,490 nm (chosen to produce a uniform Raman gain profile across the optical fibre bandwidth with $\sim 1\text{-dB}$ variation) and wavelength-division multiplexers that coupled the pump lasers into and out of the DCF. The digitizer had a bandwidth of 16 GHz and a sampling rate of 5×10^{10} samples per second. The captured 1D temporal data was sorted into a 2D matrix in the digital domain for image reconstruction.

The detection sensitivity and minimum detectable reflectivity were -148 dBm Hz^{-1} and 9.2×10^{-7} , respectively. The pixel size and field of view (horizontal \times vertical) were found to be $6.4 \mu\text{m} \times 0.3 \mu\text{m}$ and $170 \mu\text{m} \times 27 \mu\text{m}$, respectively, for Figs 2 and 4, and $16 \mu\text{m} \times 0.7 \mu\text{m}$ and $430 \mu\text{m} \times 68 \mu\text{m}$, respectively, for Fig. 3.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 December 2008; accepted 2 March 2009.

- Riehle, R. A. Jr. *Principles of Extracorporeal Shock Wave Lithotripsy* (Churchill Livingstone, 1987).
- Kodama, R. et al. Fast heating of ultrahigh-density plasma as a step towards laser fusion ignition. *Nature* **412**, 798–802 (2001).
- Epstein, I. R. & Pojman, J. A. *An Introduction to Nonlinear Chemical Dynamics* (Oxford Univ. Press, 1998).
- Petty, H. R. Spatiotemporal chemical dynamics in living cells: from information trafficking to cell physiology. *Biosystems* **83**, 217–224 (2004).
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
- Grinvald, A., Frostig, R. D., Siegel, R. M. & Bartfeld, E. High-resolution optical imaging of functional brain architecture in the awake monkey. *Proc. Natl Acad. Sci. USA* **88**, 11559–11563 (1991).
- Gridley, T. & Woychik, R. Laser surgery for mouse geneticists. *Nature Biotechnol.* **25**, 59–60 (2007).
- Yanik, M. F. et al. Functional regeneration after laser axotomy. *Nature* **432**, 822 (2004).
- Slade, S. G., Baker, R., Brockman, D. K. & Thornton, S. P. *The Complete Book of Laser Eye Surgery* (Bantam, 2002).
- Squires, T. M. & Quake, S. R. Microfluidics: fluid physics at the nanoliter scale. *Rev. Mod. Phys.* **77**, 977–1026 (2005).
- Watson, J. V. *Introduction to Flow Cytometry* (Cambridge Univ. Press, 2004).
- Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
- Fang, N., Lee, H., Sun, C. & Zhang, X. Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308**, 534–537 (2005).
- Petty, H. R. Applications of high speed microscopy in biomedical research. *Opt. Photon. News* **15**, 40–45 (2004).
- Etoh, T. G. et al. Evolution of ultra-high-speed CCD imagers. *Plasma Fusion Res.* **2**, S1021 (2007).
- Barty, A. et al. Ultrafast single-shot diffraction imaging of nanoscale dynamics. *Nature Phys.* **2**, 415–419 (2008).
- Buechler, Ch, Dong, C. Y., So, P. T. C., French, T. & Gratton, E. Time-resolved polarization imaging by pump-probe stimulated emission fluorescence microscopy. *Biophys. J.* **79**, 536–549 (2000).
- Porneala, C. & Willis, D. A. Observation of nanosecond laser-induced phase explosion in aluminum. *Appl. Phys. Lett.* **89**, 211121 (2006).
- Solli, D. R., Roper, C., Koonath, P. & Jalali, B. Optical rogue waves. *Nature* **450**, 1054–1057 (2007).

20. Diddams, S. A., Hollberg, L. & Mbele, V. Molecular fingerprinting with the resolved modes of a femtosecond laser frequency comb. *Nature* **445**, 627–630 (2007).
21. Chou, J., Boyraz, O., Solli, D. & Jalali, B. Femtosecond real-time single-shot digitizer. *Appl. Phys. Lett.* **91**, 161105 (2007).
22. Solli, D. R., Chou, J. & Jalali, B. Amplified wavelength-time transformation for real-time spectroscopy. *Nature Photon.* **2**, 48–51 (2008).
23. Goda, K., Tsia, K. K. & Jalali, B. Amplified dispersive Fourier-transform imaging for ultrafast displacement sensing and barcode reading. *Appl. Phys. Lett.* **93**, 131109 (2008).
24. Savage, N. Scientific CCD cameras. *Nature Photon* **1**, 351–353 (2007).
25. Lynch, T. F. & Zeng, A. Image intensified cameras for high-speed imaging applications that require more than just the high frame-rate. *Proc. SPIE* **5580**, 790–795 (2005).
26. Bartelt, H. O. Wavelength multiplexing for information transmission. *Opt. Commun.* **27**, 365–368 (1978).
27. Tearney, G. J., Shishkov, M. & Bouma, B. E. Spectrally encoded miniature endoscopy. *Opt. Lett.* **27**, 412–414 (2002).
28. Xiao, S. & Weiner, A. M. 2-D wavelength demultiplexer with potential for ≥ 1000 channels in the c-band. *Opt. Express* **12**, 2895–2901 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was partially supported by the US Defense Advanced Research Projects Agency and the Center for Nanoscience Innovation for Defense. We are grateful to D. R. Solli, Y. Hoshino and T. Kodama for discussions. We also thank T. Lay for creating the animated film.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.G. (goda@ee.ucla.edu).

METHODS

Optical source for the STEAM camera. The optical source used for the STEAM camera was a mode-locked femtosecond laser. A high-nonlinearity fibre was used to generate a supercontinuum with a bandwidth of about 40 nm centred around 1,590 nm. The laser was pulse-picked to reduce the repetition rate to 6.1 MHz using a gated Mach–Zehnder amplitude modulator.

2D spatial disperser. The 2D spatial disperser consisted of a pair of orthogonally oriented spatial dispersers (a virtually imaged phased array (VIPA) and diffraction grating) that produced a 2D spectral pattern which we refer to as a ‘spectral shower’. An input pulse is spatially dispersed and separated by the dispersers into many subpulses of different colours. The VIPA was essentially a tilted Fabry–Pérot cavity with a highly reflective coating ($R = 99.9\%$) on one surface except for an uncoated window area and a partially reflective coating ($R \approx 95\%$) on the other surface. A collimated laser beam was focused by a cylindrical lens ($f = 200$ mm) onto the highly reflective surface of the VIPA at a small angle ($\sim 3^\circ$) with respect to the propagation direction through the uncoated window area. The multiple interference that occurred inside the VIPA resulted in angular dispersion that is over an order of magnitude larger than that of typical diffraction gratings. The VIPA with a free spectral range (FSR) of 67 GHz and a line-width of 550 MHz dispersed the spectrum of the pulse in one direction. Wavelengths that differed by integer multiples of the FSR were spatially overlapped with each other. An orthogonally placed diffraction grating with a groove density of 1,200 lines per millimetre was used to separate this degeneracy in the other direction. It resulted in the spectrum of the incident pulse being mapped into a 2D spectral pattern in space, resembling a spectral shower.

Amplified dispersive Fourier transform. Before the amplified dispersive Fourier transformation was performed, the spectrum of the back-reflected spectral shower was pre-amplified using an L-band EDFA and filtered using a band-pass filter centred at 1,590 nm with a pass band of 15 nm. The Raman-amplified dispersive element consisted of (1) a DCF with a total group-velocity dispersion of -3.3 ns nm $^{-1}$, (2) four ~ 300 -mW continuous-wave pump lasers, positioned at regular length intervals along the DCF, with respective wavelengths of 1,470 nm, 1,480 nm, 1,490 nm and 1,480 nm, and (3) wavelength-division multiplexers that coupled the pump lasers into and out of the DCF. These pump wavelengths were chosen to produce a uniform Raman gain profile across the optical filter bandwidth with ~ 1 -dB variation. The amplified 1D temporal data stream was captured by a single-pixel photodetector with a 10-GHz bandwidth (Discovery Semiconductors) and digitized using a high-speed digital oscilloscope with a 16-GHz bandwidth and a sampling rate of 5×10^{10} samples per second (Tektronix).

Digital image processing. The purpose of the digital image processing was to sort the 1D temporal data stream into a 2D matrix for image reconstruction. This sorting step was done with knowledge of the FSR of the VIPA, which corresponds to the column length of the final image. The dispersion slope of the DCF (which results in nonlinear frequency-to-time mapping) was also included in this step to

avoid the skewing problem in the image reconstruction. In addition, the captured waveform was first digitally band-pass-filtered to remove high-frequency noise, and was normalized by the filtered background waveform corresponding to the background image. The resultant image quality was further enhanced by noise removal using the Wiener filter, image interpolation and contrast equalization.

System parameters of the STEAM camera. The field of view, pixel size, detection sensitivity and shutter speed of the STEAM camera were calculated using the experimental parameters. See the Supplementary Information for a complete description of the equations that characterize the following system parameters.

(1) Pixel size and field-of-view. On the basis of the tilt angle and FSR of the VIPA, the diffraction angle ($\sim 65^\circ$) and groove density of the diffraction grating, the centre wavelength and repetition rate of the spectral shower, and the focal length of the objective lens (~ 2 mm for imaging the calibration target and microfluidic flow, respectively (Figs 2 and 4), and ~ 5 mm for imaging the laser ablation (Fig. 3)), the pixel size and field of view (horizontal \times vertical) were found to be $6.4 \mu\text{m} \times 0.3 \mu\text{m}$ and $170 \mu\text{m} \times 27 \mu\text{m}$, respectively, for Figs 2 and 4, and $16 \mu\text{m} \times 0.7 \mu\text{m}$ and $430 \mu\text{m} \times 68 \mu\text{m}$, respectively, for Fig. 3. From these values, the number of pixels was found to be about 2,500 for all the experiments performed in this paper, and can be increased by a few orders of magnitude by increasing the optical bandwidth, the amount of temporal dispersion used in the dispersive Fourier transform process and the sampling rate of the digitizer. In addition, with the aid of virtual time gating and a parallel architecture, the number of pixels per frame can be increased without sacrificing the frame rate (Supplementary Information).

(2) Detection sensitivity. On the basis of the total noise figure (5.6 dB) and the net image gain (25 dB) of the EDFA and Raman amplifier, as well as the bandwidth (10 GHz) and noise-equivalent power ($1 \text{ pW Hz}^{-1/2}$) of the photodetector, the input referred noise was found to be -148 dBm Hz^{-1} . Furthermore, on the basis of the average power of each pulse incident on the object for the pulse duration, the minimum detectable reflectivity was found to be 9.2×10^{-7} , corresponding to a detection sensitivity (the inverse of the reflectivity) as high as 60 dB.

(3) Shutter speed. The shutter speed was found to be about 440 ps, on the basis of the pulse width of each subpulse of the spectral shower. This can be obtained from the bandwidth of each subpulse (719 MHz), assuming that the pulses are nearly transform-limited and taking dispersion in the optics into consideration.

Laser ablation experiment. The excitation pulse laser used in the laser ablation experiment was an optical parametric oscillator (OPOTEK) pumped by a Nd:YAG Q-switched laser. It generated a high-power pulse with a centre wavelength of $2.8 \mu\text{m}$ and a pulse width of 5 ns. A lens ($f = 75$ mm) produced a focused pulse with a fluence of about 20 J cm^{-2} at the surface of the sample.

Microfluidics experiment. The microfluidic system used in the microfluidics experiment consisted of a 15-cm-long hollow fibre with an inner diameter of $50 \mu\text{m}$. The metal microspheres, which had diameters ranging from 10 to $30 \mu\text{m}$ (Thermo Fisher Scientific), were suspended in water with a small amount of ethanol (for wetting the microspheres).

LETTERS

The Kondo effect in ferromagnetic atomic contacts

M. Reyes Calvo¹, Joaquín Fernández-Rossier¹, Juan José Palacios¹, David Jacob², Douglas Natelson³ & Carlos Untiedt¹

Iron, cobalt and nickel are archetypal ferromagnetic metals. In bulk, electronic conduction in these materials takes place mainly through the *s* and *p* electrons, whereas the magnetic moments are mostly in the narrow *d*-electron bands, where they tend to align. This general picture may change at the nanoscale because electrons at the surfaces of materials experience interactions that differ from those in the bulk. Here we show direct evidence for such changes: electronic transport in atomic-scale contacts of pure ferromagnets (iron, cobalt and nickel), despite their strong bulk ferromagnetism, unexpectedly reveal Kondo physics, that is, the screening of local magnetic moments by the conduction electrons below a characteristic temperature¹. The Kondo effect creates a sharp resonance at the Fermi energy, affecting the electrical properties of the system; this appears as a Fano–Kondo resonance² in the conductance characteristics as observed in other artificial nanostructures^{3–11}. The study of hundreds of contacts shows material-dependent log-normal distributions of the resonance width that arise naturally from Kondo theory¹². These resonances broaden and disappear with increasing temperature, also as in standard Kondo systems^{4–7}. Our observations, supported by calculations, imply that coordination changes can significantly modify magnetism at the nanoscale. Therefore, in addition to standard micromagnetic physics, strong electronic correlations along with atomic-scale geometry need to be considered when investigating the magnetic properties of magnetic nanostructures.

Atomic-scale contacts can be fabricated by techniques such as scanning tunnelling microscopy¹³ or the use of electromigrated break junctions (EBJs)¹⁴, where the size of a macroscopic contact between two leads is reduced until they are in contact through only a few atoms and, eventually, through only one. The conductance of metallic monatomic contacts is known to be around $2G_0$, where $G_0 = e^2/h$ is

the spin-resolved quantum of conductance¹³ (e being the elementary charge and h Planck's constant). To identify the atomic contacts, histograms are constructed from the evolution of the conductance recorded during the breaking of different contacts (Fig. 1a, b). The position of the first peak of these histograms is identified as the conductance of the monatomic contact. For iron, cobalt and nickel, the conductance is larger than $2G_0$ owing to the contribution of the *sp* and *d* orbitals to the transmission^{15–17}.

We have studied the low-temperature conductance characteristics of hundreds of atomic-scale contacts of the three transition-metal ferromagnets iron, cobalt and nickel using a home-built STM. More than the 80% of the differential conductance (dI/dV) curves at the monatomic contact show peaks or dips around zero bias such as those shown in Fig. 1c, which are very similar to those observed in STM spectroscopy of single magnetic adatoms on non-magnetic surfaces^{9–11}. Thus, as in the case of these Kondo systems, we can also fit our dI/dV curves to the sum of a flat component, g_0 , and a Fano-like resonance that typically amounts for 10% of the signal:

$$\frac{dI}{dV} = g_0 + \frac{A}{1+q^2} \frac{(q+\varepsilon)^2}{1+\varepsilon^2} \quad (1)$$

Here $\varepsilon = (eV - \varepsilon_s)/k_B T_K$ is the bias shifted with respect to the centre of the resonance, ε_s , and normalized by the natural width of the resonance, $k_B T_K$; T_K is the Kondo temperature; q is the dimensionless Fano parameter that determines the symmetry of the curve; A is the amplitude of the feature; and k_B is Boltzmann's constant.

It is clear from the dI/dV characteristics in Fig. 1c that the width of the Fano feature is different for each material. This is confirmed by statistical analysis of the data. Figure 2 shows histograms of T_K obtained from fitting our conductance curves for hundreds of different iron, cobalt and nickel monatomic contacts. Notably, these histograms

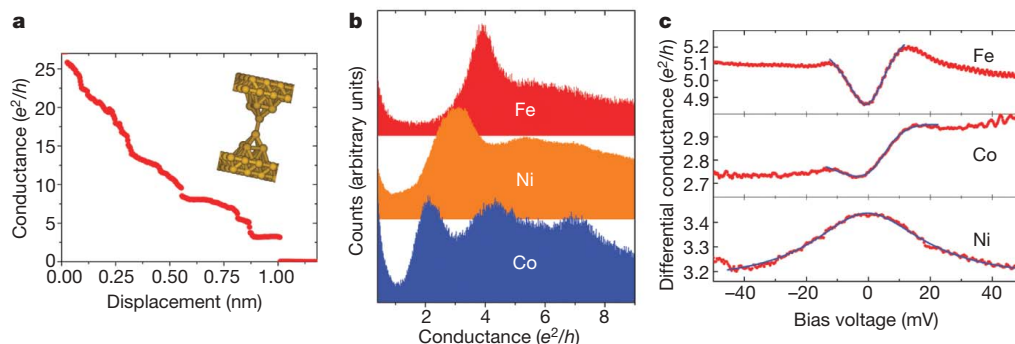


Figure 1 | Conductance of a monatomic contact. **a**, Example of a trace where we record the conductance while stretching a nickel wire using a scanning tunnelling microscope (STM) at 4.2 K. Inset, model of a monatomic contact. **b**, Conductance histograms constructed for iron, cobalt and nickel from thousands of such traces. The position of the first peak of in each histogram corresponds to the conductance of the monatomic contact. **c**, Differential

conductance curves recorded at the monatomic contact as a function of the applied voltage. A characteristic resonance appears at small bias that fits the Fano line shape. All possible symmetries are found in the spectroscopy of iron, cobalt and nickel contacts, and the width of the resonance is the main difference between the spectra of the three materials. This width is proportional to the Kondo temperature.

¹Departamento de Física Aplicada, Facultad de Ciencias, Universidad de Alicante, San Vicente del Raspeig, E-03790 Alicante, Spain. ²Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA. ³Department of Physics and Astronomy, Rice University, Houston, Texas 77005, USA.

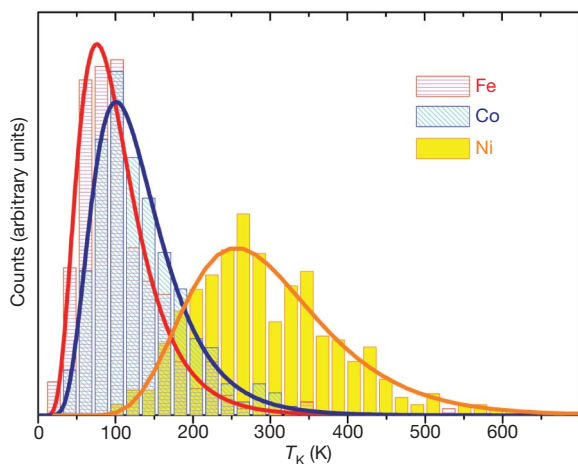


Figure 2 | Histograms of inferred Kondo temperatures for iron, cobalt and nickel. The histograms are constructed from more than 200 fittings and normalized to the total number of curves fitted. The continuous lines show the fits of the data to log-normal distributions of T_K with a different most probable value for each material.

follow a log-normal distribution; that is, the natural logarithm of T_K is normally distributed. Because many different atomic configurations result in monatomic contacts, their electronic properties, such as conductance (Fig. 1b), density of states and the associated energy scales, are expected to be normally distributed. Instead, a normal distribution of $\log T_K$ can only be understood if T_K can be expressed as the exponential of normally distributed quantities. The Kondo model¹² naturally relates T_K to the exponential of the typical energy scales in the problem.

Fitting the histograms to a log-normal distribution yields most frequent values for the resonance widths in the different materials corresponding to $T_K = 90$ K, 120 K and 280 K for iron, cobalt and nickel, respectively, following the same trend ($T_K^{\text{Fe}} < T_K^{\text{Co}} < T_K^{\text{Ni}}$) for these chemical species when deposited as adatoms on non-magnetic surfaces¹⁸. In simple terms, the Kondo temperature decreases as the size of the screened magnetic moment increases, as we change from nickel to cobalt to iron. The Kondo temperature of cobalt nano-contacts is very similar to the one reported in ref. 11 for cobalt in Cu(100) probed using a STM in the contact regime.

In addition to the statistical analysis described above, we measured the temperature evolution of the Kondo features for a single contact. To preserve the atomic stability of the junction while changing the temperature¹⁹, we used an EBJ such as the one shown in the inset of Fig. 3b. It is important to note that in this alternative implementation we observe exactly the same Fano resonances, with the same distribution of Kondo temperatures, as we obtained in the STM experiments using nickel and cobalt samples. In all the cases, we see a reduction of the amplitude of the Fano features as the temperature increases, as shown in Fig. 3a for the case of cobalt. In Fig. 3b we plot $A(T)$, as defined in equation (1), as a function of T on a logarithmic scale. The curve so obtained has a low-temperature plateau followed by a linear decay, very similar to that of quantum dots and molecules in the Kondo regime^{4–6,20}.

In summary, our atomic contacts fabricated with two different methods show the same dI/dV curves (Fig. 1c), the same chemical trends (Fig. 2) and the same temperature evolution (Fig. 3a, b) as other standard, chemically inhomogeneous Kondo systems. This indicates that the contact atom(s) in nanocontacts of iron, cobalt and nickel are in the Kondo regime. This is unexpected for two reasons. First, the Kondo effect has always been associated with chemically inhomogeneous systems containing at least two kinds of atom: those where the localized level resides and those providing the itinerant electrons. Here the same chemical species hosts both the itinerant states and the local magnetic moments. Second, in most reports on the Kondo effect, the itinerant electrons are not spin-polarized, as a large spin

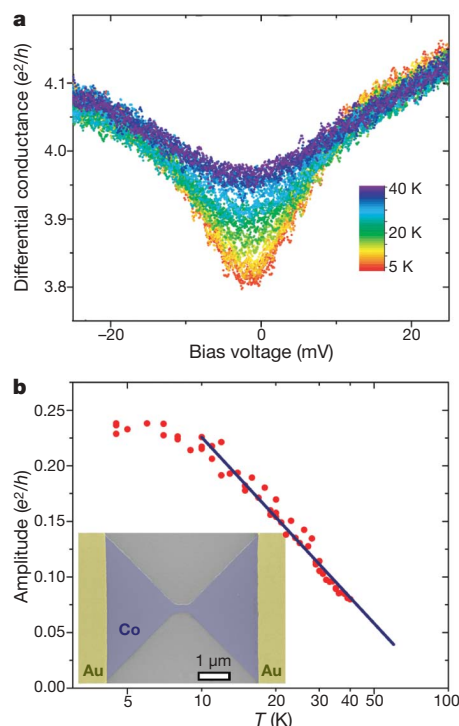


Figure 3 | Evolution of the Fano resonances with increasing temperature. **a**, Characteristics of differential conductance versus bias voltage for a cobalt atomic contact, showing how the Fano resonance disappears as the temperature is increased. **b**, The amplitude of the Kondo resonance from **a** decreases logarithmically with the temperature. Inset, an artificially coloured example of a lithographic device for the EBJ experiments (before electromigration); the cobalt junction is blue. The atomic contacts created by this method are suitable for studying the temperature dependence of the Fano resonance.

polarization is expected to destroy the effect. The observation of the Kondo effect in a quantum dot configuration such as a C_{60} molecule²¹ or a carbon nanotube²² contacted with ferromagnetic leads has been justified by the oppositely directed spin polarizations in the electrodes, in agreement with refs 23, 24. Although the situation regarding the magnetization of the leads might be similar in our system, we argue that the Kondo effect is still possible even if there is no domain wall pinned in the contact.

Being in the Kondo regime implies that the atoms at the contact must host, at least, a localized d -electron level whose magnetic moment is screened as a result of antiferromagnetic coupling to the sp conduction electrons. We can show how local-moment formation and antiferromagnetic coupling occur in nickel nanocontacts in the following way (the cases of iron and cobalt can be understood on similar grounds). The mean-field solution of the Anderson model²⁵, which describes a localized d level, with energy ϵ_d , on-site repulsion U and hybridization with the itinerant sp electrons, V_{sp-d} defines the conditions for the formation of a local moment in such a d level. The model can also be used to derive the antiferromagnetic sd exchange coupling, J_{sd}^{AF} (ref. 26), which arises from the $sp-d$ hybridization term. Here we used the local spin-density approximation (LSDA) to density functional theory and its generalization LSDA + U as the mean field to determine the formation of a local moment and its antiferromagnetic coupling to the sp carriers. We consider both monostrand chains and nanocontacts. Their smaller atomic coordination, compared with that of the bulk, results in a stronger electronic localization and a larger magnetic moment per atom (values of 1.17 Bohr magnetons, compared with 0.6 Bohr magnetons in the bulk, were obtained using LSDA). This favours the appearance of the Kondo effect. In Fig. 4a, b, we show the LSDA bands obtained for the nickel chain. Out of the six minority-spin bands crossing the Fermi level, the two degenerate E2

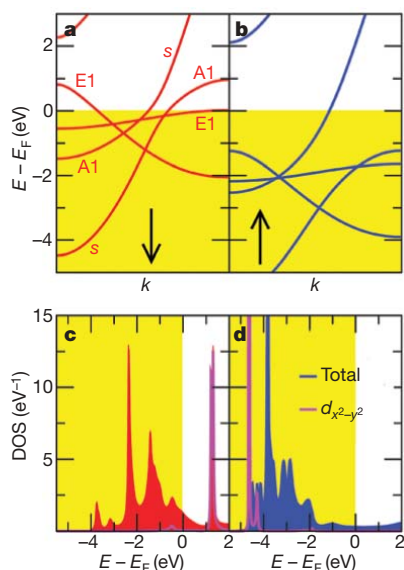


Figure 4 | Electronic structure for a nickel chain and a nickel nanocontact. **a**, LSDA minority-spin energy bands for a monostrand nickel chain, where the wavevector k runs over the first Brillouin zone. The bands are labelled according to their symmetry group: the E1 and E2 bands (d -band like, doubly degenerate and decoupled from the s band) and the A1 or $d_{3z^2-r^2}$ band (hybridized with the s band). The lattice constant is 2.09 Å. **b**, As in **a**, but for majority-spin electrons. The yellow background indicates occupied states. **c**, LSDA+U minority-spin results for the total DOS and the DOS projected on the $d_{x^2-y^2}$ orbital for a tip atom of a nickel nanocontact, with $U = 3$ eV. **d**, As in **c**, but for majority-spin electrons.

bands are the narrowest, hosting highly localized electrons. These are the bands that are less well described by the LSDA because of the inherent self-interaction problem of this approximation.

We modelled actual nanocontacts with two identical pyramids facing each other in the [001] direction. As previously done for chains²⁷, and to avoid the self-interaction problem, we computed the electronic structure of the nanocontact using LSDA+U. In Fig. 4c, d, we show the total density of states (DOS) projected on a tip atom and the DOS projected on the corresponding $d_{x^2-y^2}$ orbitals for both the minority (Fig. 4c) and the majority (Fig. 4d) spins (here we have set $U = 3$ eV). In general, solutions obtained for values of U ranging from 3 to 5 eV also show that the $d_{x^2-y^2}$ orbital, forming one of the E2 bands in chains, hosts an integer local magnetic moment. Different geometries may favour the formation of the local moment in other strongly localized orbitals. The hybridization of the $d_{x^2-y^2}$ orbital with the surrounding sp orbitals results in antiferromagnetic kinetic exchange²⁶. We estimate that $J_{sp-d}^{AF} \approx 1$ eV, taking $\varepsilon_d \approx 5$ eV, $|V_{sp-d}|^2 \approx 2$ eV and $U_{eff} \approx 8$ eV from our LSDA+U calculations, where U_{eff} is the spin splitting of the $d_{x^2-y^2}$ orbital (Supplementary Information).

We note that, in contrast to the contact atom(s), the intra-atomic $sp-d$ hybridization vanishes for bulk atoms owing to the very small anisotropy of the crystal environment. This makes the antiferromagnetic coupling J_{sp-d}^{AF} larger in the contact than in the bulk. This coupling competes with the ferromagnetic coupling J_{sp-d}^{FM} , which we estimate from the splitting of the sp band at the boundary of the Brillouin zone for chains to be ~ 0.2 eV (Fig. 4a, b). Thus, an overall antiferromagnetic coupling between d electrons and sp conduction electrons is possible in nickel nanocontacts. Additionally, the local moment, \mathbf{m} , responsible for the Kondo effect is subject to the ferromagnetic coupling J_{dd} to the neighbouring atoms, and this interaction also competes with the antiferromagnetic $sp-d$ coupling:

$$H_{\text{exch}} = \mathbf{m} \cdot \left[(J_{sp-d}^{AF} - J_{sp-d}^{FM}) \mathbf{S}_s - J_{dd} \sum_i \mathbf{m}_i \right]$$

Here the \mathbf{m}_i are the local moments of the neighbouring atoms and \mathbf{S}_s is the spin of the s electrons. In ref. 28, the coupling J_{dd} was calculated for iron, cobalt and nickel by implementing the magnetic force theorem with a LSDA ground state. This method yields values for the spin-wave dispersion of the materials that compare well with experiment²⁹, and gives $J_{dd} = 19$ meV, 15 meV and 2.7 meV for iron, cobalt and nickel, respectively. Thus, J_{dd} is significantly smaller than J_{sp-d}^{AF} .

The Kondo effect in nanocontacts is favoured by three factors. First, a local moment forms in the contact atoms because of their smaller coordination. Second, the reduced symmetry of the contact, compared with the bulk, enhances the intra-atomic contribution to the $sp-d$ hybridization and, thus, the antiferromagnetic coupling J_{sp-d}^{AF} . Third, the smaller coordination also reduces the influence of the direct ferromagnetic dd coupling with neighbouring atoms. As a result, the local moment formed in the contact is antiferromagnetically coupled to the sp itinerant electrons and results in the Kondo effect in this system, in contradiction to conventional wisdom.

METHODS SUMMARY

For the statistical results on the Kondo parameters, we used a home-made STM at 4.2 K to fabricate the contacts by indentation between two pieces of metal¹³. For the temperature-dependence measurements, the contacts were produced by the controlled electromigration at 4.2 K of 100-nm-wide junctions fabricated by electron beam lithography¹⁹. In both cases, the spectroscopic curves were obtained by the addition of an a.c. voltage with a peak-power amplitude of 1 mV and a frequency of 1 kHz to the d.c. bias voltage to allow the lock-in detection of the differential conductance.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 October 2008; accepted 16 February 2009.

- Kondo, J. Resistance minimum in dilute magnetic alloys. *Prog. Theor. Phys.* **32**, 37–49 (1964).
- Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124**, 1866–1878 (1961).
- Goldhaber-Gordon, D. *et al.* The Kondo effect in a single-electron transistor. *Nature* **391**, 156–159 (1998).
- Cronenwett, S. M., Oosterkamp, T. H. & Kouwenhoven, L. P. A tunable Kondo effect in quantum dots. *Science* **281**, 540–544 (1998).
- Park, J. *et al.* Coulomb blockade and the Kondo effect in single-atom transistors. *Nature* **417**, 722–725 (2002).
- Liang, W., Shores, M. P., Bockrath, M., Long, J. R. & Park, H. Kondo resonance in a single-molecule transistor. *Nature* **417**, 725–729 (2002).
- Nygard, J., Cobden, D. H. & Lindelof, P. E. Kondo physics in carbon nanotubes. *Nature* **408**, 342–346 (2000).
- Yu, L. H. & Natelson, D. The Kondo effect in C60 single-molecule transistors. *Nano Lett.* **4**, 79–83 (2003).
- Mandhavan, V., Chen, W., Jamneala, T., Crommie, M. F. & Wingreen, N. S. Tunneling into a single magnetic atom: spectroscopic evidence of the Kondo resonance. *Science* **280**, 567–569 (1998).
- Li, J., Schneider, W. D., Berndt, R. & Delley, B. Kondo scattering observed at a single magnetic impurity. *Phys. Rev. Lett.* **80**, 2893–2896 (1998).
- Néel, N. *et al.* Conductance and Kondo Effect in a controlled single-atom contact. *Phys. Rev. Lett.* **98**, 016801 (2006).
- Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1993).
- Agrait, N., Levy-Yeyati, A. & van Ruitenbeek, J. M. Quantum properties of atomic-sized conductors. *Phys. Rep.* **377**, 81–279 (2003).
- Park, H., Lim, A. K. L., Alivisatos, A. P., Park, J. & McEuen, P. L. Fabrication of metallic electrodes with nanometer separation by electromigration. *Appl. Phys. Lett.* **75**, 301–303 (1994).
- Untiedt, C., Dekker, D. M. T., Djukic, D. & van Ruitenbeek, J. M. Absence of magnetically induced fractional quantization in atomic contacts. *Phys. Rev. B* **69**, 081401 (2004).
- Jacob, D., Fernández-Rossier, J. & Palacios, J. J. Magnetic and orbital blocking in Ni nanocontacts. *Phys. Rev. B* **71**, 220403 (2005).
- Calvo, M. R., Caturia, M. J., Jacob, D., Untiedt, C. & Palacios, J. J. Mechanical, electrical and magnetic properties of Ni nanocontacts. *IEEE Trans. Nanotechnol.* **7**, 165–168 (2008).
- Jamneala, T., Madhavan, V., Chen, W. & Crommie, M. F. Scanning tunneling spectroscopy of transition-metal impurities at the surface of gold. *Phys. Rev. B* **61**, 9990–9993 (2000).

19. Keane, Z. K., Yu, L. H. & Natelson, D. Magnetoresistance of atomic-scale electromigrated nickel nanocontacts. *Appl. Phys. Lett.* **88**, 062514–062516 (2006).
20. van der Wiel, W. G. *et al.* The Kondo effect in the unitary limit. *Science* **289**, 2105–2108 (2000).
21. Pasupathy, A. N. *et al.* The Kondo effect in the presence of ferromagnetism. *Science* **306**, 86–89 (2004).
22. Hauptmann, J. R., Paaske, J. & Lindelof, P. E. Electric-field-controlled spin reversal in a quantum dot with ferromagnetic contacts. *Nature Phys.* **4**, 373–376 (2008).
23. Martinek, J. *et al.* Kondo effect in quantum dots coupled to ferromagnetic leads. *Phys. Rev. Lett.* **91**, 127203 (2003).
24. Martinek, J. *et al.* Kondo effect in the presence of itinerant-electron ferromagnetism studied with the numerical renormalization group method. *Phys. Rev. Lett.* **91**, 247202 (2003).
25. Anderson, P. W. Localized magnetic states in metals. *Phys. Rev.* **124**, 41–53 (1961).
26. Schrieffer, J. R. & Wolf, P. A. Relation between the Anderson and Kondo Hamiltonians. *Phys. Rev.* **149**, 491–492 (1966).
27. Wierzbowska, M., Delin, A. & Tosatti, E. Effect of electron correlations in Pd, Ni, and Co nanowires. *Phys. Rev. B* **72**, 035439 (2005).
28. Pajda, M., Kudrnovský, J., Turek, I., Drchal, V. & Bruno, P. *Ab initio* calculations of exchange interactions, spin-wave stiffness constants, and Curie temperatures of Fe, Co, and Ni. *Phys. Rev. B* **64**, 174402 (2001).
29. Mook, H. A. & Paul, D. McK. Neutron-scattering measurement of the spin-wave spectra for nickel. *Phys. Rev. Lett.* **54**, 227–230 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Tosatti, R. Aguado and J. Ferrer for discussions, G. Scott and G. Saenz-Arce for experimental support and V. Esteve for technical support. This work was partly supported by the European Union through MolSpinQIP and Spanish MEC (grants MAT2007-65487, 31099-E and CONSOLIDER CSD2007-0010). D.J. acknowledges funding by the US National Science Foundation (NSF) under grant DMR-0528969. D.N. acknowledges the support of NSF grant DMR-0347253, the David and Lucille Packard Foundation and the W. M. Keck Program in Quantum Materials.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.U. (untiedt@ua.es).

METHODS

Fabrication of atomic-scale contacts by scanning tunnelling microscopy. Two pieces of metal wire of 0.1-mm diameter were cleaned and sonicated in acetone and isopropanol. These pieces were mounted in a home-built STM. The set-up was pumped down to high vacuum and immersed in a liquid helium bath until the temperature reached 4.2 K. The two pieces of wire were brought into contact and then pulled apart until the contact was of atomic dimension and, finally, until only one atom formed the contact¹³. Strong indentation before the formation of every contact ensured the cleanliness of the atomic contacts. The histograms obtained by this technique (Fig. 1a) show similar results to the ones obtained by the mechanically controlled break junction technique, where the surfaces brought into contact are created under cryogenic conditions.

Fabrication of atomic-scale contacts using EBJs. As a first step, a small junction about 100 nm wide was fabricated from cobalt using electron beam lithography and electron beam evaporation over a silicon dioxide substrate. To make the junction suitable for contact by macroscopic probes, two gold electrodes are deposited over the edges of the junction in a second lithography step, following the procedure described in ref. 19. The samples are then placed in a probe station that was pumped down and immersed in a liquid helium bath until the sample reached a temperature close to 4.2 K. Under these conditions, the controlled electromigration process^{14,19} was performed, decreasing the size of the junction to the atomic scale.

LETTERS

The ITQ-37 mesoporous chiral zeolite

Junliang Sun¹, Charlotte Bonneau¹, Ángel Cantín², Avelino Corma², María J. Díaz-Cabañas², Manuel Moliner², Daliang Zhang¹, Mingrun Li¹ & Xiaodong Zou¹

The synthesis of crystalline molecular sieves with pore dimensions that fill the gap between microporous and mesoporous materials is a matter of fundamental and industrial interest^{1–3}. The preparation of zeolitic materials with extralarge pores and chiral frameworks would permit many new applications. Two important steps in this direction include the synthesis⁴ of ITQ-33, a stable zeolite with $18 \times 10 \times 10$ ring windows, and the synthesis⁵ of SU-32, which has an intrinsically chiral zeolite structure and where each crystal exhibits only one handedness. Here we present a germanosilicate zeolite (ITQ-37) with extralarge 30-ring windows. Its structure was determined by combining selected area electron diffraction (SAED) and powder X-ray diffraction (PXRD) in a charge-flipping algorithm⁶. The framework follows the SrSi_2 (srs) minimal net⁷ and forms two unique cavities, each of which is connected to three other cavities to form a gyroidal channel system. These cavities comprise the enantiomorphous srs net of the framework. ITQ-37 is the first chiral zeolite with one single gyroidal channel. It has the lowest framework density ($10.3 \text{ T atoms per } 1,000 \text{ \AA}^3$) of all existing 4-coordinated crystalline oxide frameworks, and the pore volume of the corresponding silica polymorph would be $0.38 \text{ cm}^3 \text{ g}^{-1}$.

Three-dimensional extra-large pore systems have been the realm of amorphous silica materials with wall structures corresponding to the primitive, diamondoid and gyroidal (G) periodic surfaces, with a significantly higher occurrence of the G-surface; only the latter exhibits chiral channels of opposite handedness on either side of the surface. The greater stability of the G-surface has been demonstrated mathematically⁸ and is understood in the chemistry of mesoporous silica materials to be a favourable configuration for reducing the hydrophobic/hydrophilic interface during crystallization. With inorganic crystalline compounds, the ANA⁹ framework remained the only zeolite following the G-surface until the discovery of UCSB-7¹⁰ with 12-ring channels. The first crystalline oxide with pores larger than 20 \AA was the germanate SU-M¹¹, with Ge in tetrahedral and octahedral coordination, which cannot be considered a zeolite. Its framework can be described as a reticulation of the G-surface. If we could direct the inorganic framework on one side of the G-surface, while leaving the other side empty, a zeolitic material with extralarge pores and channels displaying only one enantiomorph might be obtained. ITQ-37 is the first such zeolite.

The systematic exploration of the phase diagrams of germanosilicates by high-throughput techniques using three large dicationic organic structure directing agents (SDA1, SDA2 and SDA3; Supplementary Fig. 1) was the key for discovering the new zeolite ITQ-37 (see Methods). Whereas SDA1 and SDA3 gave beta-type materials or clathrasils, SDA2 produced ITQ-24¹² or a mixture of ITQ-24 and ITQ-37 (Supplementary Tables 1 and 2). SDA2 is a bulky diquaternary ammonium molecule (Fig. 1). We analysed the different synthesis parameters to find the synthesis conditions leading to pure ITQ-37 (see Methods and Supplementary Figs 2, 3 and 4). The chemical formula of ITQ-37 is $[(\text{C}_{22}\text{N}_2\text{H}_{40})_{10.5}(\text{H}_2\text{O})_x][\text{Ge}_{80}\text{Si}_{112}\text{O}_{400}\text{H}_{32}\text{F}_{20}]$,

as obtained by combining the structure refinement and elemental analysis. SDA2 has been shown to be intact in the final product.

The germanosilicate ITQ-37 crystallizes in a cubic chiral space group ($P4_132$ or $P4_332$, $a = 26.5126(3) \text{ \AA}$), as determined by combining SAED and PXRD (see Methods Summary). The high degree of overlapping reflections ($>94\%$ exact overlapping with $d > 1.2 \text{ \AA}$; Fig. 2) and the peak broadening due to the small crystal sizes ($70\text{--}200 \text{ nm}$, Supplementary Fig. 4a) posed great challenges for the structure determination. Several complex zeolite structures were recently determined by combining transmission electron microscopy and PXRD^{13–15}. Crystal structure factor phases obtained from high-resolution transmission electron microscopy (HRTEM) images were used to facilitate the structure solution from PXRD, using the FOCUS program¹⁶ (which uses crystal chemical information) or the Superflip program (based on a charge-flipping algorithm)¹⁷.

However, these methods could not be applied to ITQ-37 because ITQ-37 is much more electron beam sensitive than other zeolites, permitting only the observation of the approximate pore structure from HRTEM images (Supplementary Fig. 5). We developed a new strategy of combining SAED and PXRD to solve the structure of ITQ-37, because SAED extends to a higher resolution than do HRTEM images (Supplementary Fig. 5). Intensities of nearly 80% of the total unique reflections to a 3.3 \AA resolution were extracted from the SAED patterns along the $[100]$, $[110]$, $[111]$ and $[120]$ directions using the program ELD¹⁸; those were used for the pre-repartitioning of overlapping reflections in PXRD. Of the 16 overlapping reflection groups with $2\theta < 26^\circ$, the intensities of reflections within ten groups were significantly improved by the pre-repartitioning step and are close to the structure factor intensities of the final structure. The charge-flipping algorithm was then applied for the structure determination. A hundred runs of charge-flipping iterations were tested using the Superflip program¹⁷. Random initial phases were used for each run, and electron density histogram matching was applied during the iterations for further repartitioning of the overlapping reflections⁶. The ten electron density maps with the best R -values ($<26\%$) were

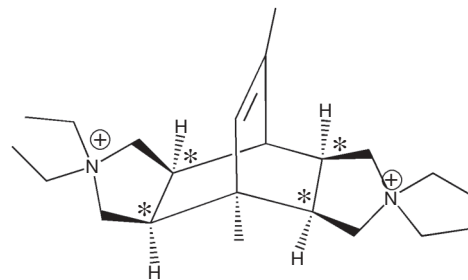


Figure 1 | Structure of SDA2 used for synthesizing the ITQ-37 zeolite. SDA2 contains four chiral centres (marked with asterisks) in a meso conformation, making the overall molecule achiral.

¹Structural Chemistry and Berzelii Centre, EXSELENT on Porous Materials, Stockholm University, SE-106 91, Stockholm, Sweden. ²Instituto de Tecnología Química (UPV-CSIC), Av. Naranjos s/n, E-46022 Valencia, Spain.

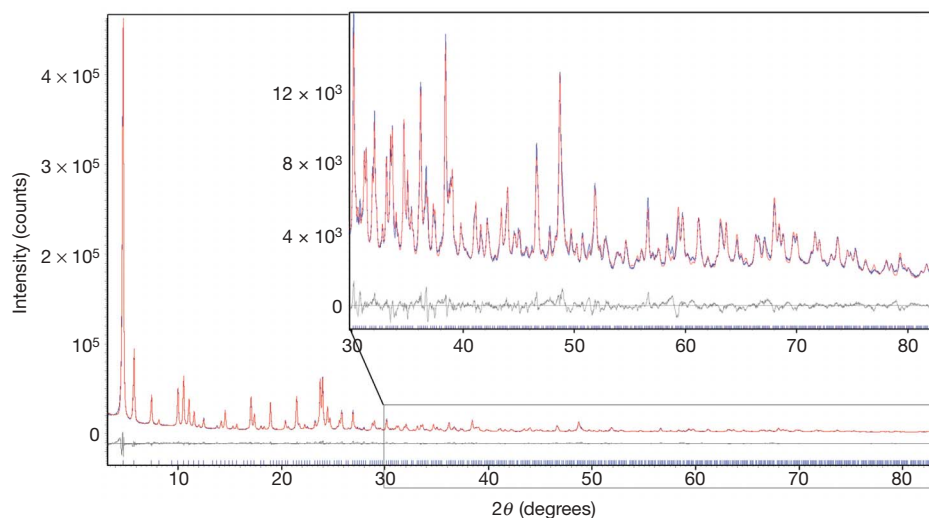


Figure 2 | Observed (blue), calculated (red) and difference (black) PXRD profiles for the Rietveld refinement of the as-synthesized ITQ-37 ($\lambda = 1.5406 \text{ \AA}$). The higher angle data ($2\theta = 30\text{--}83^\circ$) has been scaled up (inset) to show the good fit between the observed and the calculated patterns.

examined and six of them showed similar pore features and framework structures.

Ten unique T atoms (in TO_4 tetrahedra, where T is an Si or Ge atom) and 18 of the 19 unique oxygen atoms could be automatically located from the best electron density map (lowest R -value, Fig. 3) by the EDMA program. One missing oxygen between T4 and T5 could easily be identified and thus added manually. The resulting structure is a three-dimensional framework, where Si and Ge atoms are tetrahedrally (T) coordinated with oxygen. All T–O and T–T distances are reasonable except for one ($d_{\text{T4-T5}} = 3.6 \text{ \AA}$). We note that the correct model could also be obtained without the pre-repartitioning but the convergence was slower and thus decreased the chance of success. The final framework structure was refined by Rietveld refinement with $R_p = 0.0381$, $R_{wp} = 0.0503$ and $R_{exp} = 0.0115$ (p, profile; wp, weighted profile; exp, expected) (Fig. 2, Methods Summary and Supplementary Table 3). All the bond distances and angles are reasonable in the refined structure. Eight of the ten unique T sites (T2 to T9) are shared by Si and Ge and the other two (T1 and T10) are occupied mainly by Ge, each coordinated to one hydroxyl group (Supplementary Table 4).

ITQ-37 has a very open framework consisting of one unique *lau* cage [4^26^4] and two unique double 4-rings [4^6]; both are recurring motifs in zeolites. Double 4-rings have one (D4R1) or two (D4R2)

terminal hydroxyl groups (Fig. 4a). It is complicated to describe the framework of ITQ-37 using these units individually, so we chose a single tertiary building unit: $\text{T}_{44}\text{O}_{145}(\text{OH})_7$ (where $T = \text{Si, Ge}$). The tertiary building unit has the point group C3 and is made up of one D4R1, three *lau* cages and three D4R2s (Fig. 4b). Each tertiary building unit is linked to three neighbouring building units, by sharing a common D4R2 and half of a 6-ring of the *lau* unit. There are eight such tertiary building units per unit cell, which lie on the nodes of a three-coordinated srs net (Fig. 4b)¹⁹. The srs net is compatible with a gyroidal theme and usually occurs as an enantiomorphous pair separated by the G-surface. This is also the case in ITQ-37. The gyroidal framework on one side of the G-surface generates a gyroidal channel system on the other side. The framework and the channel system have opposite handedness and both follow the srs net as an enantiomorphous pair (Fig. 4c). The channel system consists of two unique large cavities at the 4a and 4b positions (with coordinates $3/8, 3/8, 3/8$ and $7/8, 7/8, 7/8$ and their symmetry equivalents). Each large cavity communicates with three others through windows of thirty TO_4 tetrahedra (30-rings) with an asymmetric opening of $4.3 \times 19.3 \text{ \AA}$ (Fig. 4c and Supplementary Fig. 6a), assuming the van der Waals diameter for oxygen of 2.7 \AA . The D4R1s, D4R2s and *lau* cages individually follow the srs, srs-e (lcv) and srs-a nets, respectively^{11,19}. All of these are chiral and of the same handedness as the

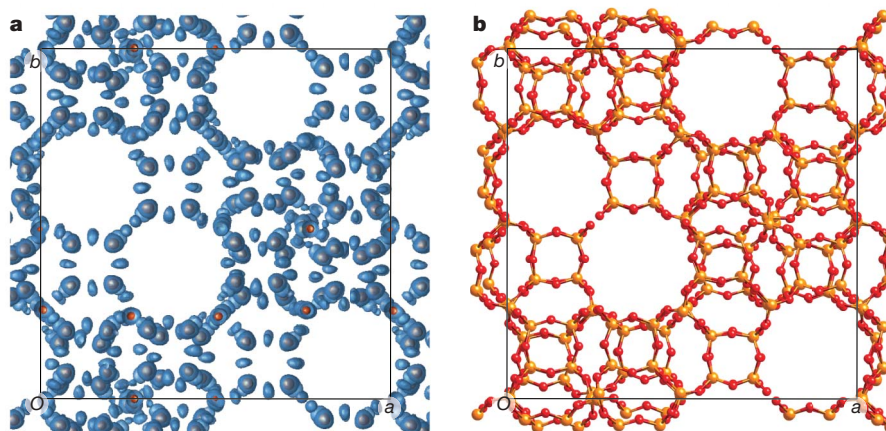


Figure 3 | Electron density map derived by the charge-flipping algorithm and the obtained structure model, both are viewed along the c axis. **a**, The electron density map is represented by two iso-surfaces at $2.7e \text{ \AA}^{-3}$ (in red) and $0.7e \text{ \AA}^{-3}$ (in blue), respectively (the maximum and minimum electron

densities are $5.5e \text{ \AA}^{-3}$ and $-0.7e \text{ \AA}^{-3}$, respectively). All T (Si and Ge) atoms and 18 of the 19 O atoms could be directly located from the map. **b**, The framework structure model of ITQ-37 deduced from **a**. The T atoms are in yellow and oxygen atoms are in red.

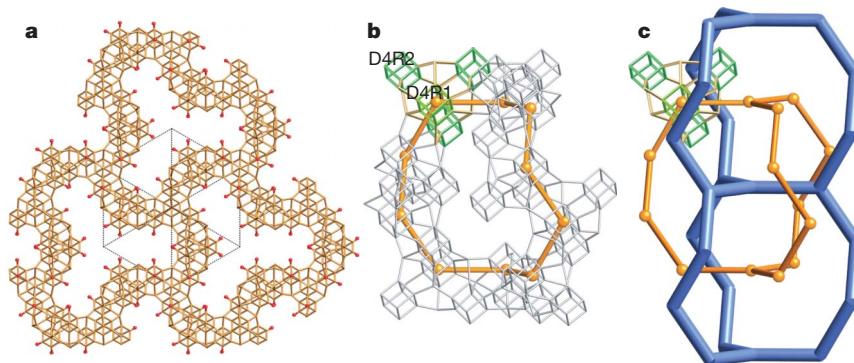


Figure 4 | The framework and corresponding nets of ITQ-37. **a**, A slice (15.3 Å thick) viewed down the [111] direction. Only the T–T connections and the terminal hydroxyl groups are shown. All double 4-rings have the same orientation. **b**, The 30-ring built from ten tertiary building units. One of them is highlighted (*lau* cage in orange and double 4-rings in green). The

centres of the tertiary building units fall on the nodes of one srs net (in orange). **c**, The large cavity defined by three 30-rings. The centres of the large cavities fall on the nodes of another srs net (in blue) that represents the gyroidal channel system.

overall framework. The structure of the channel system (Supplementary Fig. 6b and c and Supplementary Movie 1) agrees with that observed in the HRTEM images (Supplementary Fig. 5c and d).

The ^{19}F magic-angle spinning nuclear magnetic resonance (MAS NMR) spectrum (Supplementary Fig. 7) shows a single resonance band at -10 p.p.m., characteristic for fluoride anions entrapped in germanium-rich double 4-rings. The ^{29}Si MAS NMR spectrum (Supplementary Fig. 8) shows two small peaks at -94 and -96 p.p.m., which correspond to structural defects in different crystallographic positions. The close proximity of protons to these Si sites is confirmed by the enhancement of these two peaks in the cross-polarization ^1H MAS NMR spectrum.

ITQ-37 is the first chiral zeolite with a single gyroidal channel. It has the lowest framework density of all existing 4-coordinated oxide frameworks, 10.3 T atoms per $1,000 \text{ Å}^3$. The tile representation (see Methods Summary) is a valuable tool for discerning the complex framework and channel system. The tertiary building unit is constructed of five different tiles: two corresponding to the double 4-rings, one to the *lau* cage (t-*lau*) and two interfacial tiles (6^3 , t-*kah*) between the t-*lau* tiles (Fig. 5a and b). These tiles are commonly found in zeolites. The channel system is constructed of three tiles: two large tiles with similar volumes corresponding to the large cavities (Fig. 5c) and an interfacial tile (Supplementary Fig. 9). The tiling clearly demonstrates the opposite handedness of the channel system and the framework in ITQ-37 (Fig. 5c). It also offers a direct appreciation of the large accessible volume (Fig. 5d and Supplementary Movies 2 and 3). The BET (Brunauer–Emmett–Teller physical adsorption model) surface area of ITQ-37 is $690 \text{ m}^2 \text{ g}^{-1}$, with a micropore volume of $0.29 \text{ cm}^3 \text{ g}^{-1}$, as determined from the N_2 isotherm (Supplementary Fig. 10). We note that these values would correspond to a BET surface area of $900 \text{ m}^2 \text{ g}^{-1}$ and a micropore volume of $0.38 \text{ cm}^3 \text{ g}^{-1}$ for the silica polymorph of ITQ-37, which are far beyond the values previously reported for any zeolite structure.

The zeolite ITQ-37 shows a good thermal stability (Supplementary Fig. 11). A pelletized sample of ITQ-37 calcined at 813 K remained stable during two weeks while stored at room temperature in a moisture-free environment. Nevertheless, exposure to 90% humidity at room temperature led to a 20% and 50% loss of crystallinity after 10 h and 24 h, respectively. An Al-containing ITQ-37 sample ((Si+Ge)/Al = 70) was synthesized (Methods). The Al is tetrahedrally coordinated (Supplementary Fig. 12) and presents Brønsted acidity upon pyridine adsorption. Acetalization reactions require catalysts of mild acidities, and acetals are of interest for fine chemical industries^{20–22}, so we performed acetalization of aldehydes of different molecular sizes with triethyl orthoformate using ITQ-37 as the catalyst to show the benefits of the large pores. We compared the catalytic results with those obtained on zeolite beta (Si/Al = 50) with a similar crystallite size (Supplementary

Fig. 4 and Supplementary Table 5). For the smaller aldehyde (heptanal) that could diffuse in the pores of beta, both materials gave a similar initial activity. However for a larger aldehyde (diphenylacetaldehyde), the initial activity of ITQ-37 was almost three times that of beta. This indicates the presence of larger exploitable pores in ITQ-37 compared to those of beta. Furthermore, the selectivity to acetal at high conversion is much better for ITQ-37 with the bulkier aldehyde.

The discovery of the ITQ-37 zeolite has not only shown that zeolites can reach pore size dimension approaching the mesoporous range, but also provided new insight towards targeting chiral crystalline frameworks with extralarge pores. The success of this study arises from a combination of examining large non-surfactant molecules as

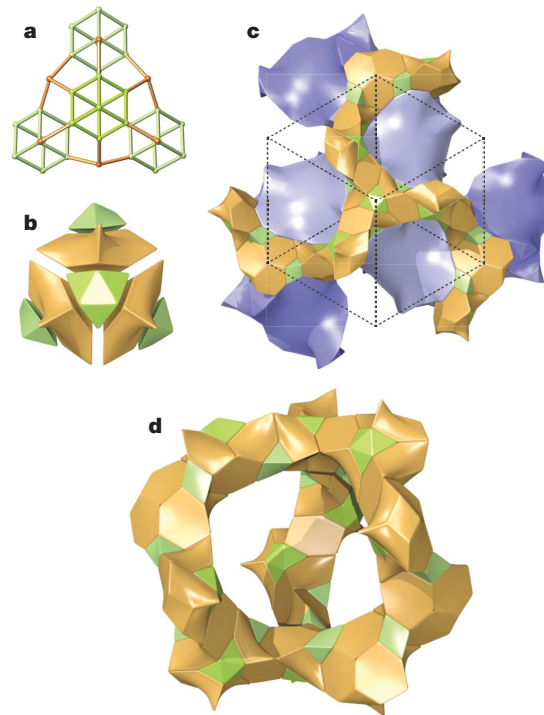


Figure 5 | Tiling of ITQ-37. **a**, The tertiary building unit $\text{T}_{44}\text{O}_{145}(\text{OH})_7$ built from one D4R1, three D4R2s [4^6] and *lau* cages [$4^2 6^4$]. **b**, Tiles of the tertiary building unit in **a**. The interfacial tiles [6^3] are omitted for clarity and hydroxy-bearing T atoms were omitted for the tiling computation. **c**, Tiling of the framework (green and orange) and the channel system (blue), showing the large pore volume compared to the framework wall. The framework and channel systems have opposite chirality (left-handedness for the framework and right-handedness for the channel system). **d**, The srs large cavity defined by three 30-rings.

organic structure directing agents and the use of TEM and PXRD techniques in a new structure solution method.

METHODS SUMMARY

TEM characterization. TEM work was performed on JEM3010 electron microscopes under low-dose conditions (Supplementary Fig. 5). The 4- and 6-fold symmetries in SAED patterns indicated a cubic cell. The unit cell was determined from SAED patterns and further refined by PXRD to be $a = 26.5126(3)$ Å. The reflection condition $h00: h = 4n$ indicates that the space group is $P4_132$ or $P4_332$ (chiral with opposite handedness). The product is expected to be a racemic mixture owing to the achiral nature of the SDA2. The projection symmetries along the $[100]$ and $[111]$ directions are $p4gm$ and $p3m1$, respectively, determined from the phases extracted from Fourier transforms of HRTEM images using CRISP²³, which further confirmed the space group. The projected potential maps clearly show the pore structure that is consistent with the structure model (Supplementary Figs 5 and 6).

Structure determination. PXRD data were collected on a PANalytical X'Pert Pro. The structure was solved by charge-flipping algorithm combining SAED intensities and PXRD using the Superflip program¹⁷. The final structure model was refined by Rietveld refinement using TOPAS²⁴ with soft restraints for the T–O bond distances considering the Si/Ge occupancies. All T positions were refined with mixed occupancies of Si and Ge and a fixed overall Si/Ge ratio of 1.40, obtained from the elemental analysis. Although the organic structure directing agents were intact in the final product, as confirmed by the elemental analysis, they could not be located owing to their partial occupancies and lower symmetry. Instead, 12 unique carbon and four oxygen atoms were added at random positions inside the pores and refined subsequently. The PXRD pattern of the calcined sample also agrees with the structure model, but the diffraction peaks were too broad at the high angle region to perform a good refinement. This may be due to the collapse of the hydroxyl groups¹⁵.

Topological analysis. The embedding of the tetrahedral net and tiling visualization were performed by Systre²⁵ and 3dt²⁶, respectively; both are part of the GAVROG project. The tiling data was computed with TOPOS²⁷.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 September 2008; accepted 6 March 2009.

- Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417**, 813–821 (2002).
- Corma, A. State of the art and future challenges of zeolites as catalysis. *J. Catal.* **216**, 298–312 (2003).
- Férey, G. Materials science: the simplicity of complexity—rational design of giant pores. *Science* **291**, 994–995 (2001).
- Corma, A., Díaz-Cabañas, M. J., Jorda, J. L., Martínez, C. & Moliner, M. High-throughput synthesis and catalytic properties of a molecular sieve with 18- and 10-member rings. *Nature* **443**, 842–845 (2006).
- Tang, L. Q. et al. A zeolite family with chiral and achiral structures built from the same building layer. *Nature Mater.* **7**, 381–385 (2008).
- Baerlocher, Ch., McCusker, L. B. & Palatinus, L. Charge flipping combined with histogram matching to solve complex crystal structures from powder diffraction data. *Z. Kristallogr.* **222**, 47–53 (2007).
- Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Three-periodic nets and tilings: regular and quasiregular nets. *Acta Crystallogr. A* **59**, 22–27 (2003).
- Schröder, G. E., Fogden, A. & Hyde, S. T. Bicontinuous geometries and molecular self-assembly: comparison of local curvature and global packing variation in genus-three cubic, tetragonal and rhombohedral surfaces. *Eur. Phys. J. B* **54**, 509–524 (2006).

- Taylor, W. H. The structure of analcite ($\text{NaAlSi}_2\text{O}_6 \cdot \text{H}_2\text{O}$). *Z. Kristallogr.* **74**, 1–19 (1930).
- Gier, T. E., Bu, X., Feng, P. & Stucky, G. D. Synthesis and organization of zeolite-like materials with three-dimensional helical pores. *Nature* **395**, 154–157 (1998).
- Zou, X., Conradsson, T., Klingstedt, M., Dadachov, M. S. & O'Keeffe, M. A mesoporous germanium oxide with crystalline pore walls and its chiral derivative. *Nature* **437**, 716–719 (2005).
- Cantín, Á., Corma, A., Díaz-Cabañas, M. J., Jorda, J. L. & Moliner, M. Rational design and HT techniques allow the synthesis of new IWR zeolite polymorphs. *J. Am. Chem. Soc.* **128**, 4216–4217 (2006).
- Gramm, F. et al. Complex zeolite structure solved by combining powder diffraction and electron microscopy. *Nature* **444**, 79–81 (2006).
- Baerlocher, Ch. et al. Structure of the polycrystalline zeolite catalyst IM-5 solved by enhanced charge flipping. *Science* **315**, 1113–1116 (2007).
- Baerlocher, Ch. et al. Ordered silicon vacancies in the framework structure of the zeolite catalyst SSZ-74. *Nature Mater.* **7**, 631–635 (2008).
- Grosse-Kunstleve, R. W., McCusker, L. B. & Baerlocher, Ch. Powder diffraction data and crystal chemical information combined in an automated structure determination procedure for zeolites. *J. Appl. Cryst.* **30**, 985–995 (1997).
- Palatinus, L. & Chapuis, G. Superflip—a computer program for the solution of crystal structures by charge flipping in arbitrary dimensions. *J. Appl. Cryst.* **40**, 786–790 (2007).
- Zou, X. D., Sukharev, Y. & Hovmöller, S. Quantitative measurement of intensities from electron diffraction patterns for structure determination—new features in the program system ELD. *Ultramicroscopy* **52**, 436–444 (1993).
- O'Keeffe, M., Peskov, M. A., Ramsden, S. J. & Yaghi, O. M. The Reticular Chemistry Structure Resource (RCSR) database of, and symbols for, crystal nets. *Acc. Chem. Res.* **41**, 1782–1798 (2008).
- Climent, M. J., Corma, A. & Velty, A. Zeolites for the production of fine chemicals. Synthesis of the fructose fragrance. *J. Catal.* **196**, 345–351 (2000).
- Climent, M. J., Corma, A. & Velty, A. Design of a solid catalyst for the synthesis of a molecule with an orange blossom scent. *Green Chem.* **4**, 565–569 (2002).
- Climent, M. J., Corma, A. & Velty, A. Synthesis of hyacinth, vanilla and orange blossom fragrances. The benefit of using zeolites and delaminated zeolites as catalysts. *Appl. Catal. Gen.* **263**, 155–161 (2004).
- Hovmöller, S. CRISP: Crystallographic image processing on a personal computer. *Ultramicroscopy* **41**, 121–135 (1992).
- Young, R. A. *The Rietveld Method* 1–39 (IUCr Book Series, Oxford Univ. Press, 1993).
- Delgado-Friedrichs, O. & O'Keeffe, M. Identification and symmetry computation for crystal nets. *Acta Crystallogr. A* **59**, 351–360 (2003).
- Delgado-Friedrichs, O. Data structures and algorithms for tilings. I. *Theor. Comput. Sci.* **303**, 431–445 (2003).
- Blatov, V. A. Multipurpose crystallochemical analysis with the program package TOPOS. *IUCr Comp. Commun. Newsl.* **7**, 4–38 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This project is supported by the CICYT (Project MAT 2006-14274-C02-01 and Prometeo 2008 GV), the Swedish Research Council (VR) and the Swedish Governmental Agency for Innovation Systems (VINNOVA). J.S. and C.B. are supported by post-doctoral grants from the Carl-Trygger and Wenner-Gren foundations respectively. M.M. thanks ITQ for a scholarship.

Author Contributions D.Z. and M.L. carried out the TEM work. J.S. solved and refined the structures. C.B. did the topological analysis. M.M., M.J.D.-C. and A.C. carried out the zeolite synthesis work. A.C. synthesized the organic structure directing agents. J.S., C.B., A.C. and X.Z. wrote and corrected the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.C. (acorma@itq.upv.es) or X.Z. (zou@struc.su.se).

METHODS

Synthesis of SDA1 and SDA2. The starting materials for the synthesis of SDA1 and SDA2, were bicyclo[2.2.1]oct-7ene-2,3,5,6-tetracarboxylic dianhydride, and 4,6-dimethyl- α -pyrone and maleic anhydride, respectively.

To synthesise the Diels–Alder adduct from 4,6-dimethyl- α -pyrone and maleic anhydride, a toluene solution (500 ml) of 4,6-dimethyl- α -pyrone (161 mmol) and maleic anhydride (322 mmol) was refluxed for 5 days. After cooling, the resulting precipitate was filtered and washed with hexane to give the corresponding bicyclodanhydride (87%).

To aminate the bicyclodanhydrides, we dissolved the bicyclodanhydride products (140 mmol) in ethylamine solution (70% in H₂O) (400 ml) and refluxed for 3.5 days. After cooling, the solvent was removed in vacuum, providing the desired diimides in quantitative yields.

To reduce the diimides, we slowly added the corresponding diimide (49 mmol) to a suspension of LiAlH₄ (244 mmol) in anhydrous THF (300 ml) under N₂ and at 0 °C. When the addition was finished the mixture was refluxed for 5 h and stirred at room temperature overnight. The reaction was then quenched by addition of H₂O (10 ml), 15% aqueous solution of NaOH (10 ml) and distilled H₂O (10 ml). After 30 min stirring at room temperature the solution was filtered, partially concentrated under vacuum and then extracted with dichloromethane. The combined organic extracts were washed with brine, dried and concentrated to dryness to yield the corresponding amines (66%).

To alkylate the diamine, we added iodomethane (642 mmol) to a solution of diamine (52 mmol) in 70 ml of methanol. The mixture was stirred at room temperature for 72 h and methyl iodide (642 mmol) was added and stirred for an additional 72 h. The final organic dications were then concentrated under vacuum and precipitated by addition of diethyl ether. The precipitate was filtered under vacuum, yielding 20.8 g (89%) of the diquaternary ammonium as diiodide salt.

Synthesis of SDA3. The starting materials for the synthesis of SDA3 were *N*-methylmaleimide and benzene.

For the cycloaddition of *N*-methylmaleimide and benzene, a solution of *N*-methylmaleimide (108 mmol) in a mixture of benzene (300 ml), acetophenone (30 ml) and acetone (84 ml) was distributed in ten Pyrex tubes. Before the photochemical reaction, N₂ was flowed through the solutions for 15 min and then the solutions were irradiated with a high-pressure mercury lamp (200 nm < λ < 90 nm) while stirring for 48 h. The resulting precipitate was filtered under vacuum to yield the desired diimide (40%).

To reduce the diimide, we slowly added the corresponding diimide (49 mmol) to a suspension of LiAlH₄ (244 mmol) in anhydrous THF (300 ml) under N₂ and at 0 °C. The mixture was then refluxed for 5 h and stirred at room temperature overnight. The reaction was quenched by addition of H₂O (10 ml), 15% aqueous solution of NaOH (10 ml) and distilled H₂O (10 ml). After 30 min stirring at room temperature (about 25 °C) the solution was filtered, partially concentrated under vacuum and extracted with dichloromethane. The combined organic extracts were washed with brine, dried and concentrated to dryness to yield the corresponding diamine (70%).

To alkylate the diamine, we added a solution of the diamine (33.5 mmol) in methanol (85 ml) to CH₃I (1.7 mol). The mixture was stirred for 7 days at room temperature. The mixture was concentrated under vacuum to yield the desired diammonium salt (75%).

Synthesis of ITQ-37. Synthesis gels were prepared using an automated system, composed of a robotic arm for vial handling and weighing of solids, a stirring station for gel homogenization and evaporation, a liquid dosing station equipped with pumps and an analytical balance.

The typical automated synthesis procedure is as follows: GeO₂ and boric acid (99.5%, Aldrich) or alumina (74.6%, Condea) were dissolved in the structure directing agent hydroxide solution. Then colloidal silica (Ludox AS-40, Aldrich) was added, and finally, a solution of NH₄F (98%, Aldrich).

The synthesis gel was transferred to Teflon vials (3 ml), which were finally inserted in a 15-well multiautoclave. Crystallization was carried out at 175 °C under static conditions. After filtration, washing and drying, the samples were characterized by PXRD using a multi-sample Phillips X'Pert diffractometer employing Cu K α radiation.

The compositions of two typical synthesis gels to obtain ITQ-37 are as follows: 0.5 SiO₂:0.5 GeO₂:0.25 SDA2(OH)₂:0.50 NH₄F:3 H₂O, without aluminium, and 0.5 SiO₂:0.5 GeO₂:0.01 Al₂O₃:0.25 SDA2(OH)₂:0.50 NH₄F:3 H₂O, with aluminium. Crystallization was carried out at 175 °C over 24 h under static conditions.

Catalytic experiments. Activation of 100 mg of the catalyst was performed *in situ* by heating the solid at 110 °C under vacuum for 3 h. After this time, the system was left to cool down to room temperature and then a solution of the carbonyl compound (Aldrich) (3 mmol) and triethyl orthoformate (98%, Aldrich) (11 mmol) in tetrachloromethane (Panreac, 25 ml as solvent) was poured onto the activated catalyst. The resulting suspension was magnetically stirred at reflux temperature. Aliquots were analysed at different reaction times by means of gas chromatography (HP-5 column), while products were identified by mass spectroscopy. The response factors of the different compounds were determined to calculate accurately the conversion and selectivity of the process.

LETTERS

Greenhouse-gas emission targets for limiting global warming to 2 °C

Malte Meinshausen¹, Nicolai Meinshausen², William Hare^{1,3}, Sarah C. B. Raper⁴, Katja Frieler¹, Reto Knutti⁵, David J. Frame^{6,7} & Myles R. Allen⁷

More than 100 countries have adopted a global warming limit of 2 °C or below (relative to pre-industrial levels) as a guiding principle for mitigation efforts to reduce climate change risks, impacts and damages^{1,2}. However, the greenhouse gas (GHG) emissions corresponding to a specified maximum warming are poorly known owing to uncertainties in the carbon cycle and the climate response. Here we provide a comprehensive probabilistic analysis aimed at quantifying GHG emission budgets for the 2000–50 period that would limit warming throughout the twenty-first century to below 2 °C, based on a combination of published distributions of climate system properties and observational constraints. We show that, for the chosen class of emission scenarios, both cumulative emissions up to 2050 and emission levels in 2050 are robust indicators of the probability that twenty-first century warming will not exceed 2 °C relative to pre-industrial temperatures. Limiting cumulative CO₂ emissions over 2000–50 to 1,000 Gt CO₂ yields a 25% probability of warming exceeding 2 °C—and a limit of 1,440 Gt CO₂ yields a 50% probability—given a representative estimate of the distribution of climate system properties. As known 2000–06 CO₂ emissions³ were ~234 Gt CO₂, less than half the proven economically recoverable oil, gas and coal reserves^{4–6} can still be emitted up to 2050 to achieve such a goal. Recent G8 Communiqués⁷ envisage halved global GHG emissions by 2050, for which we estimate a 12–45% probability of exceeding 2 °C—assuming 1990 as emission base year and a range of published climate sensitivity distributions. Emissions levels in 2020 are a less robust indicator, but for the scenarios considered, the probability of exceeding 2 °C rises to 53–87% if global GHG emissions are still more than 25% above 2000 levels in 2020.

Determining probabilistic climate change for future emission scenarios is challenging, as it requires a synthesis of uncertainties along the cause–effect chain from emissions to temperatures; for example, uncertainties in the carbon cycle⁸, radiative forcing and climate responses. Uncertainties in future climate projections can be quantified by constraining climate model parameters to reproduce historical observations of temperature⁹, ocean heat uptake¹⁰ and independent estimates of radiative forcing. By focusing on emission budgets (the cumulative emissions to stay below a certain warming level) and their probabilistic implications for the climate, we build on pioneering mitigation studies^{11,12}. Previous probabilistic studies—while sometimes based on more complex models—either considered uncertainties only in a few forcing components¹³, applied relatively simple likelihood estimators ignoring the correlation structure of the observational errors¹⁴ or constrained only model parameters like climate sensitivity rather than allowed emissions.

Using a reduced complexity coupled carbon cycle–climate model^{15,16}, we constrain future climate projections, building on the Fourth IPCC Assessment Report (AR4) and more recent research. In particular, multiple uncertainties in the historical temperature observations⁹ are treated separately for the first time; new ocean heat uptake estimates are incorporated¹⁰; a constraint on changes in effective climate sensitivity is introduced; and the most recent radiative forcing uncertainty estimates for individual forcing agents are considered¹⁷.

The data constraints provide us with likelihood estimates for the chosen 82-dimensional space of climate response, gas-cycle and radiative forcing parameters (Supplementary Fig. 3). We chose a Bayesian approach, but also obtain ‘frequentist’ confidence intervals for climate sensitivity (68% interval, 2.3–4.5 °C; 90%, 2.1–7.1 °C), which is in approximate agreement with the recent AR4 estimates. Given the inherent subjectivity of Bayesian priors, we chose priors for climate sensitivity such that we obtain marginal posteriors identical to 19 published climate sensitivity distributions (Fig. 1a). These distributions are not all independent and not equally likely, and cannot be formally combined¹⁸. They are used here simply to represent the wide variety of modelling approaches, observational data and likelihood derivations used in previous studies, whose implications for an emission budget have not been analysed before. For illustrative purposes, we chose the climate sensitivity distribution of ref. 19 with a uniform prior in transient climate response (TCR, defined as the global-mean temperature change which occurs at the time of CO₂ doubling for the specific case of a 1% yr^{–1} increase of CO₂) as our default. This distribution closely resembles the AR4 estimate (best estimate, 3 °C; likely range, 2.0–4.5 °C) (Supplementary Information).

Maximal warming under low emission scenarios is more closely related to the TCR than to the climate sensitivity¹⁹. The distribution of the TCR of our climate model for the illustrative default is slightly lower than derived within another model set-up¹⁹, but within the range of results of previous studies (Fig. 1b), and encompasses the range arising from emulations by coupled atmosphere–ocean general circulation models¹⁶ (AOGCMs) (Fig. 1c).

Representing current knowledge on future carbon-cycle responses is difficult, and might be best encapsulated in the wide range of results from the process-based C4MIP carbon-cycle models⁸. We emulate these C4MIP models individually by calibrating 18 parameters in our carbon-cycle model¹⁶, and combine these settings with the other gas cycles, radiative forcing and climate response parameter uncertainties gained from our historical constraining.

Additional challenges arise in estimating the maximum temperature change resulting from a certain amount of cumulative emissions. The analysis needs to be based on a multitude of emission pathways with realistic multi-gas characteristics^{20,21}, as well as varying

¹Potsdam Institute for Climate Impact Research, Telegrafenberg, 14412 Potsdam, Germany. ²Department of Statistics, University of Oxford, South Parks Road, Oxford OX1 3TG, UK. ³Climate Analytics, Telegrafenberg, 14412 Potsdam, Germany. ⁴Centre for Air Transport and the Environment, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK. ⁵Institute for Atmospheric and Climate Science, ETH Zurich, 8092 Zurich, Switzerland. ⁶Smith School of Enterprise and the Environment, University of Oxford, Oxford OX1 2BQ, UK. ⁷Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

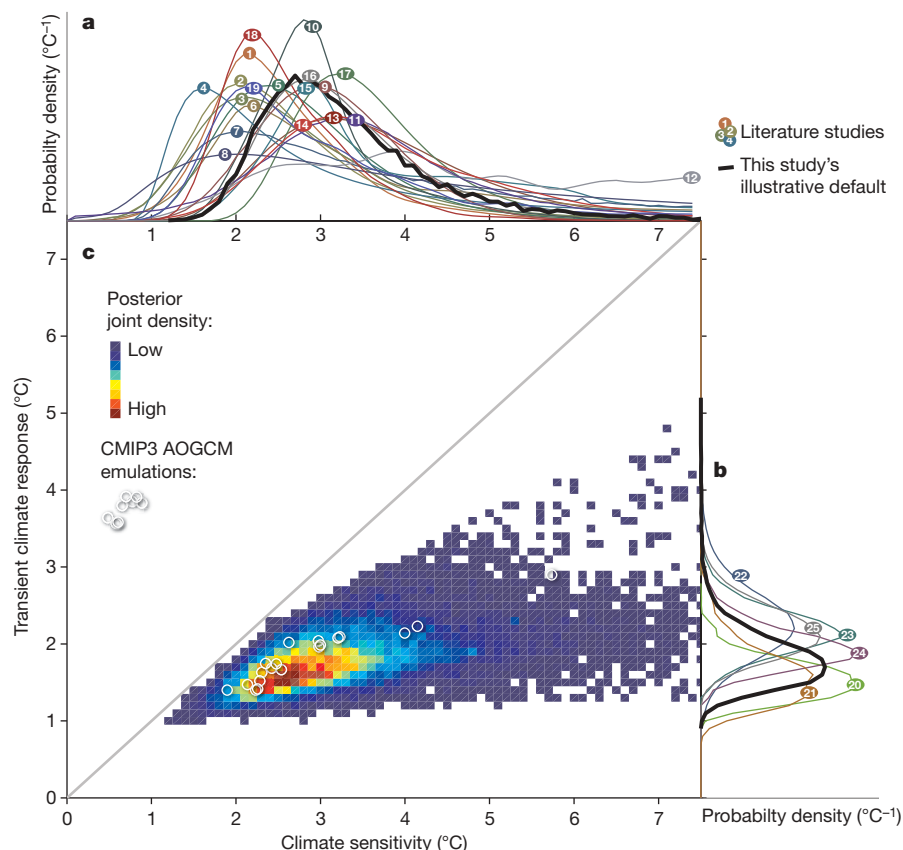


Figure 1 | Joint and marginal probability distributions of climate sensitivity and transient climate response. **a**, Marginal probability density functions (PDFs) of climate sensitivity; **b**, marginal PDFs of transient climate response (TCR); **c**, posterior joint distribution constraining model parameters to historical temperatures, ocean heat uptake and radiative forcing under our

representative illustrative priors. For comparison, TCR and climate sensitivities are shown in **c** for model versions that yield a close emulation of 19 CMIP3 AOGCMs (white circles)¹⁶. Data sources for curves 1–25 are given in Supplementary Information.

shapes over time. AOGCM results for multi-gas mitigation scenarios were not available for assessment in the IPCC AR4 Working Group I Report²². Consequently, IPCC AR4 Working Group III²³ provided equilibrium warming estimates corresponding to 2100 radiative forcing levels for some multi-gas mitigation scenarios, using simplified regressions (Supplementary Fig. 6). Thus, 15 years after the first pioneering mitigation studies^{11,12}, there is still an important gap in the literature relating emission budgets for lower emission profiles to the probability of exceeding maximal warming levels; a gap that this study intends to fill.

We compute time-evolving distributions of radiative forcing and surface air temperature implications for the set of 26 IPCC SRES²¹ and 20 EMF-21 scenarios²⁰ shown in Fig. 2a and b. We complement these with 948 multi-gas equal quantile walk emission pathways²⁴ that share—by design—similar multi-gas characteristics (Supplementary Fig. 5) but represent a wide variety of plausible shapes, ranging from early moderate reductions to later peaking and rapidly declining emissions towards near-zero emissions (Supplementary Information). Whereas Fig. 2e shows a standard plot of global-mean temperature versus time for two sample scenarios, Fig. 2f highlights the strong correlation between maximum warming and cumulative emissions. The fraction of climate model runs above 2 °C (dashed line in Fig. 2f) is then our estimate for the probability of exceeding 2 °C for an individual scenario (as indicated by the dots in Fig. 3a). We focus here on 2 °C relative to pre-industrial levels, as such a warming limit has gained increasing prominence in science and policy circles as a goal to prevent dangerous climate change²⁵. We recognize that 2 °C cannot be regarded as a ‘safe level’, and that (for example) small island states and least developed countries are calling for warming to be limited to 1.5 °C (Supplementary Information).

We chose the twenty-first century as our time horizon, as this time frame is sufficiently long to determine which emission scenarios will probably lead to a global surface warming below 2 °C. Under these scenarios, temperatures have stabilized or peaked by 2100, while warming continues under higher scenarios.

For our illustrative distribution of climate system properties, we find that the probability of exceeding 2 °C can be limited to below 25% (50%) by keeping 2000–49 cumulative CO₂ emissions from fossil sources and land use change to below 1,000 (1,440) Gt CO₂ (Fig. 3a and Table 1). If we resample model parameters to reproduce 18 published climate sensitivity distributions, we find a 10–42% probability of exceeding 2 °C for such a budget of 1,000 Gt CO₂. If the acceptable exceedance probability were only 20%, this would require an emission budget of 890 Gt CO₂ or lower (illustrative default). Given that around 234 Gt CO₂ were emitted between 2000 and 2006 and assuming constant rates of 36.3 Gt CO₂ yr^{−1} (ref. 3) thereafter, we would exhaust the CO₂ emission budget by 2024, 2027 or 2039, depending on the probability accepted for exceeding 2 °C (respectively 20%, 25% or 50%).

To contrast observationally constrained probabilistic projections against current AOGCM and carbon-cycle models, we ran each emission scenario with all permutations of 19 CMIP3²⁶ AOGCM and 10 C4MIP carbon-cycle model emulations¹⁶. The allowed emissions are similar to the lower part of the range spanned by the observationally constrained distributions, suggesting that the current AOGCMs do not substantially over- or underestimate future climate change compared to the values obtained using a model constrained by observations, although no probability statement can be derived from the proportion of runs exceeding 2 °C (black dashed line in Fig. 3a). Using an independent approach focusing on CO₂ alone, Allen *et al.*²⁷

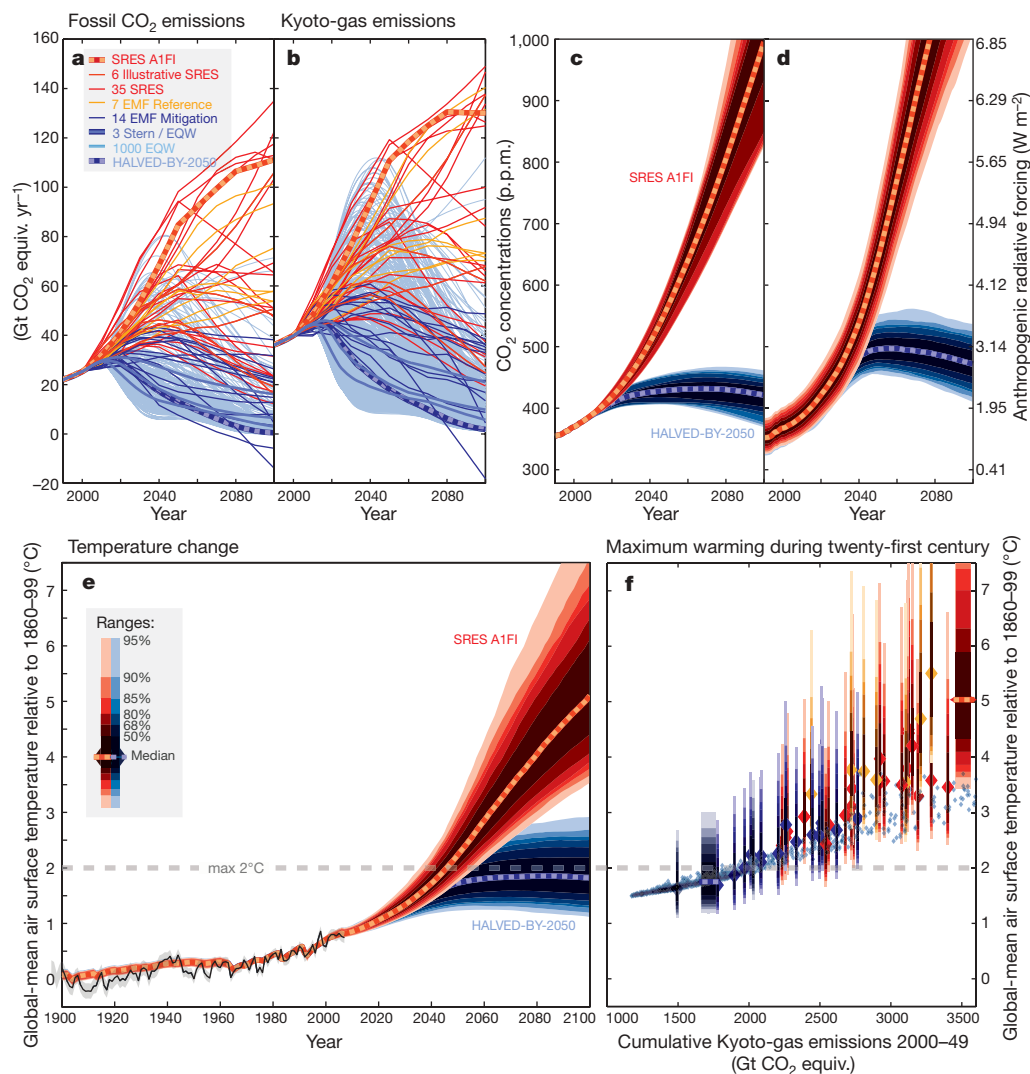


Figure 2 | Emissions, concentrations and twenty-first century global-mean temperatures. **a**, Fossil CO₂ emissions for IPCC SRES²¹, EMF-21²⁰ scenarios and a selection of equal quantile walk²⁴ (EQW) pathways analysed here; **b**, GHGs, as controlled under the Kyoto Protocol; **c**, median projections and uncertainties based on our illustrative default case for atmospheric CO₂ concentrations for the high SRES A1FI²¹ and the low HALVED-BY-2050³⁰

scenario, which halves 1990 global Kyoto-gas emissions by 2050; **d**, total anthropogenic radiative forcing; **e**, surface air global-mean temperature; **f**, maximum temperature during the twenty-first century versus cumulative Kyoto-gas emissions for 2000–49. Colour range shown in **e** also applies to **c**, **d** and **f**.

find that a range of 2,050–2,100 Gt CO₂ emissions from year 2000 onwards cause a most likely CO₂-induced warming of 2 °C: in the idealized scenarios they consider that meet this criterion, between 1,550 and 1,950 Gt CO₂ are emitted over the years 2000 to 2049.

We explored the consequences of burning all proven fossil fuel reserves (the fraction of fossil fuel resources that is economically recoverable with current technologies and prices: Fig. 3b and Methods). We derived a mid-estimate of 2,800 Gt CO₂ emissions from the literature, with an 80%-uncertainty range of 2,541 to 3,089 Gt CO₂. Emitting the carbon from all proven fossil fuel reserves would therefore vastly exceed the allowable CO₂ emission budget for staying below 2 °C.

Although the dominant anthropogenic warming contribution is from CO₂ emissions, non-CO₂ GHG emissions add to the risk of exceeding warming thresholds during the twenty-first century. We estimate that the so-called non-CO₂ ‘Kyoto gases’ (methane, nitrous oxide, hydrofluorocarbons, perfluorocarbons and SF₆) will constitute roughly one-third of total CO₂ equivalent (CO₂ equiv.) emissions based on 100-yr global warming potentials²⁸ over the 2000–49 period. Under our illustrative distribution for climate system properties, and taking into account all positive and negative forcing agents as provided

by Table 2.12 in AR4¹⁷, the cumulative Kyoto-gas emission budget for 2000–50 is 1,500 (2,000) Gt CO₂ equiv., if the probability of exceeding 2 °C is to be limited to approximately 25% (50%) (Table 1).

For the lower scenarios, Kyoto-gas emissions in the year 2050 are a remarkably good indicator for probabilities of exceeding 2 °C, because for these scenarios (with emissions in 2050 below ~30 Gt CO₂ equiv.), radiative forcing peaks around 2050 and temperature soon thereafter. This is indicated by the narrow spread of individual scenarios’ exceedance probabilities for similar 2050 Kyoto-gas emissions, as shown in Supplementary Fig. 1b. If emissions in 2050 are half 1990 levels, we estimate a 12–45% probability of exceeding 2 °C (Table 1) under these scenarios.

Emissions in 2020 are a less robust indicator of maximum warming (note the wide vertical spread of individual scenario dots in Supplementary Fig. 1c)—even if restricted to this class of relatively smooth emission pathways. However, the probability of exceeding 2 °C rises to 75% if 2020 emissions are not lower than 50 Gt CO₂ equiv. (25% above 2000). Given the substantial recent increase in fossil CO₂ emissions (20% between 2000 and 2006)³, policies to reduce global emissions are needed urgently if the ‘below 2 °C’ target²⁹ is to remain achievable.

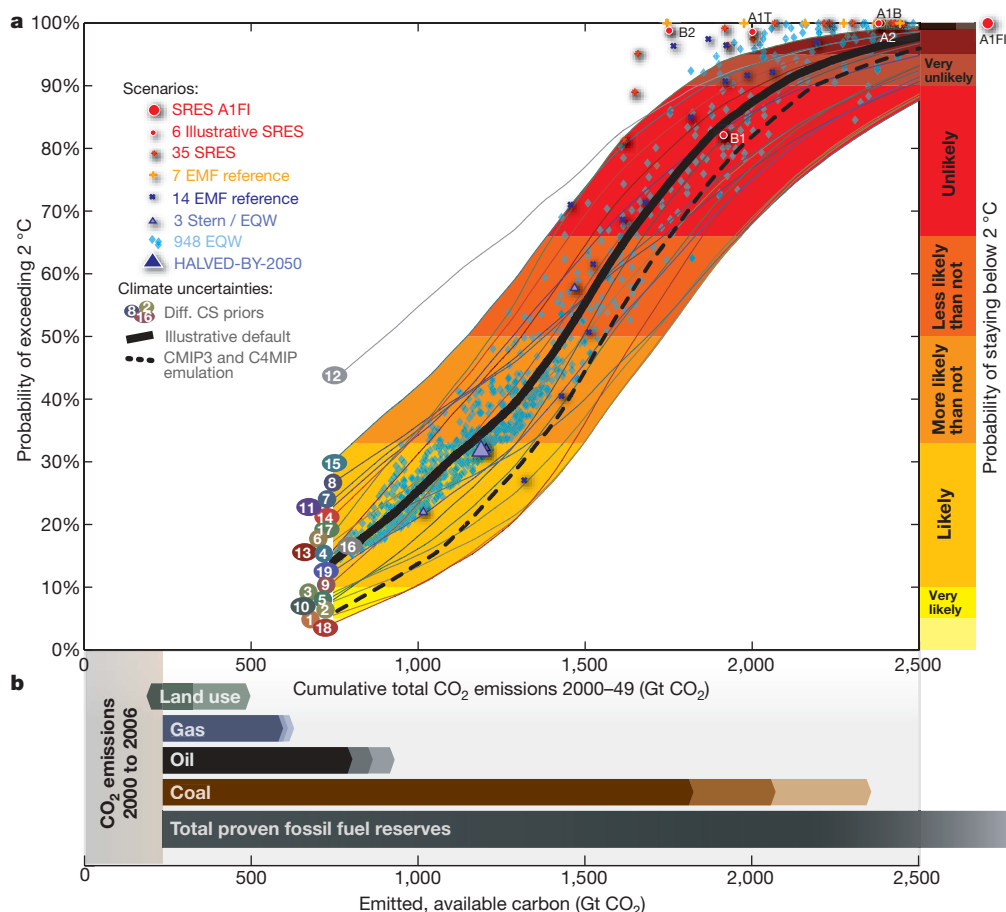


Figure 3 | The probability of exceeding 2 °C warming versus CO₂ emitted in the first half of the twenty-first century. a, Individual scenarios' probabilities of exceeding 2 °C for our illustrative default (dots; for example, for SRES B1, A2, Stern and other scenarios shown in Fig. 2) and smoothed (local linear regression smoother) probabilities for all climate sensitivity distributions (numbered lines, see Supplementary Information for data sources). The proportion of CMIP3 AOGCMs²⁶ and C4MIP carbon-cycle⁸

model emulations exceeding 2 °C is shown as black dashed line. Coloured areas denote the range of probabilities (right) of staying below 2 °C in AR4 terminology, with the extreme upper distribution (12) being omitted.

b, Total CO₂ emissions already emitted³ between 2000 and 2006 (grey area) and those that could arise from burning available fossil fuel reserves, and from land use activities between 2006 and 2049 (median and 80% ranges, Methods).

Table 1 | Probabilities of exceeding 2 °C

Indicator	Emissions	Probability of exceeding 2 °C*	
		Range	Illustrative default case†
Cumulative total CO ₂ emission 2000–49	886 Gt CO ₂	8–37%	20%
	1,000 Gt CO ₂	10–42%	25%
	1,158 Gt CO ₂	16–51%	33%
	1,437 Gt CO ₂	29–70%	50%
Cumulative Kyoto-gas emissions 2000–49	1,356 Gt CO ₂ equiv.	8–37%	20%
	1,500 Gt CO ₂ equiv.	10–43%	26%
	1,678 Gt CO ₂ equiv.	15–51%	33%
	2,000 Gt CO ₂ equiv.	29–70%	50%
2050 Kyoto-gas emissions	10 Gt CO ₂ equiv. yr ⁻¹	6–32%	16%
	(Halved 1990) 18 Gt CO ₂ equiv. yr ⁻¹	12–45%	29%
	(Halved 2000) 20 Gt CO ₂ equiv. yr ⁻¹	15–49%	32%
	36 Gt CO ₂ equiv. yr ⁻¹	39–82%	64%
2020 Kyoto-gas emissions	30 Gt CO ₂ equiv. yr ⁻¹	(8–38%)†	(21%)†
	35 Gt CO ₂ equiv. yr ⁻¹	(13–46%)†	(29%)†
	40 Gt CO ₂ equiv. yr ⁻¹	(19–56%)†	(37%)†
	50 Gt CO ₂ equiv. yr ⁻¹	(53–87%)†	(74%)†

* Range across all priors reflecting the various climate sensitivity distributions with the exception of line 12 in Fig. 3a.

† Note that 2020 Kyoto-gas emissions are, from a physical perspective, a less robust indicator for maximal twenty-first century warming with a wide scenario-to-scenario spread (Supplementary Fig. 1c).

‡ Prior chosen to match posterior of ref. 19 with uniform priors on the TCR.

METHODS SUMMARY

To relate emissions of GHGs, tropospheric ozone precursors and aerosols to gas-cycle and climate system responses, we employ MAGICC 6.0¹⁶, a reduced complexity coupled climate–carbon cycle model used in past IPCC assessment reports for emulating AOGCMs. Out of more than 400 parameters, we vary 9 climate response parameters (one of which is climate sensitivity), 33 gas-cycle

and global radiative forcing parameters (not including 18 carbon-cycle parameters, which are calibrated separately¹⁶ to C4MIP carbon-cycle models⁸), and 40 scaling factors determining the regional 4 box pattern of key forcings (Supplementary Table 1). Other parameters are set to default values¹⁶.

To constrain the parameters, we use observational data of surface air temperature⁹ in 4 spatial grid boxes from 1850 to 2006, the linear trend in ocean heat content changes¹⁰ from 1961 to 2003 and year 2005 radiative forcing estimates

for 18 forcing agents¹⁷, in addition to a constraint on the twenty-first century change of effective climate sensitivity derived from AOGCM CMIP3 emulations¹⁶. With a Metropolis-Hastings Markov chain Monte Carlo approach, based on a large ensemble ($>3 \times 10^6$) of parameter sets using 45 parallel Markov chains with 75,000 runs each, we estimate the posterior distribution of different MAGICC parameters. Estimated likelihoods take into account observational uncertainty and climate variability from various AOGCM control runs, HadCM3 being the default.

For forward projections with the model, we combine, at random, 600 sets of the 82 historically constrained parameters with one of 10 carbon-cycle calibrations. We supplemented 26 multi-gas IPCC SRES²¹ and 20 EMF-21 reference and mitigation scenarios²⁰ by 948 equal quantile walk multi-gas pathways²⁴. The proven fossil fuel reserve estimates for natural gas, oil and coal were compiled from various sources^{4,5} by combining the reserve estimates with net calorific values and emission factors (and their 95% uncertainty ranges) according to IPCC 2006 guidelines⁶ (Supplementary Information).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 September 2008; accepted 25 March 2009.

- Pachauri, R. K. & Reisinger, A. (eds) *Climate Change 2007: Synthesis Report* (Intergovernmental Panel on Climate Change, Cambridge, UK, 2007).
- Council of the European Union. *Presidency Conclusions – Brussels, 22/23 March 2005* (European Commission, 2005).
- Canadell, J. G. *et al.* Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl Acad. Sci. USA* **104**, 18866–18870 (2007).
- Clarke, A. W. & Trinnaman, J. A. (eds) *2007 Survey of Energy Resources* (World Energy Council, 2007).
- Rempe, H. Schmidt, S. & Schwarz-Schampera, U. *Reserves, Resources and Availability of Energy Resources 2006* (German Federal Institute for Geosciences and Natural Resources, 2007).
- Eggelston, H. S., Buendia, L., Miwa, K., Ngara, T. & Tanabe, K. (eds) *2006 Guidelines for National Greenhouse Gas Inventories* (IPCC National Greenhouse Gas Inventories Programme, Hayama, Japan, 2006).
- G8. *Hokkaido Toyako Summit Leaders Declaration* (G8, 2008); available at (http://www.mofa.go.jp/policy/economy/summit/2008/doc/doc080714__en.html).
- Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
- Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.* **111**, D12106, doi:10.1029/2005JD006548 (2006).
- Domingues, C. M. *et al.* Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature* **453**, 1090–1093 (2008).
- Enting, I. G., Wigley, T. M. L. & Heimann, M. *Future Emissions and Concentrations of Carbon Dioxide: Key Ocean/Atmosphere/Land Analyses* (Research technical paper no. 31, CSIRO Division of Atmospheric Research, 1994).
- Wigley, T. M. L., Richels, R. & Edmonds, J. A. Economic and environmental choices in the stabilization of atmospheric CO₂ concentrations. *Nature* **379**, 240–243 (1996).
- Forest, C. E., Stone, P. H., Sokolov, A., Allen, M. R. & Webster, M. D. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295**, 113–117 (2002).
- Knutti, R., Stocker, T. F., Joos, F. & Plattner, G. K. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**, 719–723 (2002).
- Wigley, T. M. L. & Raper, S. C. B. Interpretation of high projections for global-mean warming. *Science* **293**, 451–454 (2001).
- Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. Emulating IPCC AR4 atmosphere-ocean and carbon cycle models for projecting global-mean, hemispheric and land/ocean temperatures: MAGICC 6.0. *Atmos. Chem. Phys. Discuss.* **8**, 6153–6272 (2008).
- Forster, P. *et al.* in *IPCC Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge Univ. Press, 2007).
- Knutti, R. & Hegerl, G. C. The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nature Geosci.* **1**, 735–743 (2008).
- Frame, D. J., Stone, D. A., Stott, P. A. & Allen, M. R. Alternatives to stabilization scenarios. *Geophys. Res. Lett.* **33**, L14707, doi:10.1029/2006GL025801 (2006).
- Van Vuuren, D. P. *et al.* Temperature increase of 21st century mitigation scenarios. *Proc. Natl Acad. Sci. USA* **105**, 15258–15262 (2008).
- Nakicenovic, N. & Swart, R. *IPCC Special Report on Emissions Scenarios* (Cambridge Univ. Press, 2000).
- Solomon, S. *et al.* (eds) *IPCC Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).
- Metz, B., Davidson, O. R., Bosch, P. R., Dave, R. & Meyer, L. A. (eds) *IPCC Climate Change 2007: Mitigation* (Cambridge Univ. Press, 2007).
- Meinshausen, M. *et al.* Multi-gas emission pathways to meet climate targets. *Clim. Change* **75**, 151–194 (2006).
- Schellnhuber, J. S., Cramer, W., Nakicenovic, N., Wigley, T. M. L. & Yohe, G. *Avoiding Dangerous Climate Change* (Cambridge Univ. Press, 2006).
- Meehl, G. A., Covey, C., McAvaney, B., Latif, M. & Stouffer, R. J. Overview of coupled model intercomparison project. *Bull. Am. Meteorol. Soc.* **86**, 89–93 (2005).
- Allen, M. R. *et al.* Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* doi:10.1038/nature08019 (this issue).
- Houghton, J. T. *et al.* (eds) *IPCC Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).
- den Elzen, M. G. J. & Meinshausen, M. Meeting the EU 2°C climate target: global and regional emission implications. *Clim. Policy* **6**, 545–564 (2006).
- Watkins, K. *et al.* *Fighting Climate Change: Human Solidarity in a Divided World* (Human Development Report 2007/2008, Palgrave Macmillan, 2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Wigley, M. Schaeffer, K. Briffa, R. Schofield, T. S., von Deimling, J. Nabel, J. Rogelj, V. Huber and A. Fischlin for discussions and comments on earlier manuscripts and our code, J. Gregory for AOGCM diagnostics, D. Giebitz-Rheinbay and B. Kriemann for IT support and the EMF-21 modelling groups for providing their emission scenarios. M.M. thanks DAAD and the German Ministry of Environment for financial support. We acknowledge the modelling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model data set. Support of this data set is provided by the Office of Science, US Department of Energy.

Author Contributions M.M. and N.M. designed the research with input from W.H., R.K. and M.A. M.M. performed the climate modelling, N.M. the statistical analysis, W.H. the compilation of fossil fuel reserve estimates; all authors contributed to writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Accompanying datasets are available at www.primap.org. Correspondence and requests for materials should be addressed to M.M. (malte.meinshausen@pik-potsdam.de).

METHODS

Coupled carbon cycle–climate model. We use a reduced complexity coupled carbon cycle climate model (MAGICC 6.0), requiring (hemispheric) emissions of GHGs, aerosols, and tropospheric ozone precursors as inputs for calculating atmospheric concentrations, radiative forcings, surface air temperatures, and ocean heat uptake. MAGICC is able to closely emulate both CMIP3³⁶ AOGCMs and C4MIP⁸ carbon-cycle models, and has been used extensively in past IPCC assessment reports¹⁶. We use MAGICC 6.0 here both for future climate projections based on historical constraints and for emulating more complex AOGCMs or carbon-cycle models. The model contains many parameters whose values are uncertain. We looked at the impact of 82 parameters on model behaviour, which are summarized in the vector Θ .

Observational constraints. As one set of observational constraints, we use yearly averaged temperatures in our four grid boxes (Northern and Southern Hemisphere Land and Ocean) as provided in ref. 9 for the years 1850–2006. We arrange those measurements in a 628-dimensional vector \mathbf{T} . The respective space-time dependency of the errors is obtained from ref. 9. We use the full-length control runs of all AOGCMs runs available at PCMDI (<http://www.pcmdi.llnl.gov/>, as of mid-2007) to assess internal variability. We project the 628-dimensional vector of temperature observations into a low-dimensional subspace. We choose m so that 99.95% of the MAGICC variance is preserved and find that an eight-dimensional subspace is sufficient but findings are insensitive to this choice. We then find the $m \times 628$ -dimensional matrix P_m , which corresponds to the projection of \mathbf{T} into the space spanned by the first m PCA components. The likelihood is finally based on the m -dimensional vector $\mathbf{T}_m = P_m \mathbf{T}$ instead of the 628-dimensional vector \mathbf{T} . We now assume that the internal variability of \mathbf{T}_m has a Gaussian distribution and estimate the $m \times m$ -dimensional covariance matrix Σ_m from the data set as $P_m \Sigma P_m^T$, where Σ is the previously derived covariance matrix of the observations (including internal variability and measurement errors).

Ocean heat uptake is only considered via its linear trend Z_1 of $+0.3721$ (± 0.0698) $10^{22} \text{ J yr}^{-1}$ for the heat content trend over 1961 to 2003 up to 700 m depth¹⁰. See Supplementary Fig. 2 for the match between the constrained model results and the observational data³¹ as well as more recent results¹⁰.

Radiative forcing estimates as listed in ref. 17 (Table 2.12 therein) provide an additional set of 17 constraints Z_2, \dots, Z_{18} (Supplementary Table 2). The error of 14 of these radiative forcing estimates is assumed to have a Gaussian distribution. The remaining 3 observational constraints, however, exhibit skewness, which we model by a distribution we call here ‘skewed normal’ (Supplementary Information). All radiative forcing uncertainties are assumed to be independent.

Given that MAGICC 6.0 has substantially more freedom to change the effective climate sensitivity over time¹⁶ than what is observed from AOGCM diagnostics, we introduce another constraint Z_{19} . This constraint limits the ratio of the twenty-first century change in effective climate sensitivity, expressed by the ratio of average effective climate sensitivities in the periods 2050–2100 and 1950–2000. Based on AOGCM CMIP3 model emulations¹⁶, we derive a distribution with a median at 1.23 (with a 90% range between 1.06 to 1.51) under the SRES A1B scenario.

Likelihood estimation. To calculate the likelihood, the observations are split into the projected temperature observations \mathbf{T}_m and the remaining observational constraints Z_1, \dots, Z_{19} . Let f be the density of temperature observations under a given parameter setting Θ , taking into account both the measurement errors and internal climate variability. Let h_k , $k = 1, \dots, 19$, be the density functions of the remaining observational constraints. Under independence of Z_1, \dots, Z_{19} and \mathbf{T} , the likelihood $L(\Theta)$ of model parameters Θ is given by:

$$L(\Theta) = f(\mathbf{T}_m | \Theta) \prod_{k=1}^{19} h_k(Z_k | \Theta)$$

We follow mostly a Bayesian approach. A prior distribution π over the parameter vector Θ is specified in various ways as discussed further below, see Supplementary Table 1 for prior assumption on key parameters. Given the a priori assumption, we are able to specify the posterior distribution $g(\Theta)$ of the parameters as proportional to the product of the likelihood $L(\Theta)$ and the prior $\pi(\Theta)$.

Sensitivity to the chosen prior and a comparison with frequentist inference are discussed further below. For frequentist inference, we work directly with the likelihood.

Model sampling. To draw models from the posterior distribution $g(\Theta)$, we use a Markov chain Monte Carlo approach and a standard Metropolis–Hastings algorithm with adaptive step sizes to attain an average acceptance rate of 60%. 45 Markov chains are run in parallel for 75,000 iterations each. Adjusting for a burn-in time of 20,000 iterations, and retaining only every 30th model, to decrease dependence between successive models, a total of 82,500 models are drawn from the posterior distribution. For probabilistic forecasts, 600 models with maximal spacing in this set of 82,500 models are retained and combined randomly with one of the 10 parameter sets used for emulating individual C4MIP carbon-cycle models¹⁶.

Representation of climate sensitivity distributions. Apart from the frequentist likelihood confidence intervals, we represent the wide range of literature studies on Bayesian climate sensitivity distributions^{19,32–41}. Specifically, we change the prior for climate sensitivity such that a match between our posterior PDF of climate sensitivity and the posterior distribution in the considered studies is achieved.

Fossil fuel reserves. Our median estimates of proven recoverable fossil fuel reserves are based on ref. 42, with the exception of the non-conventional oil reserves which are taken as the median between ref. 43 and ref. 44. Potential emissions are estimated using IPCC 2006 default net calorific values and carbon content emission factors⁶ (Table 1.2 and Table 1.3 therein).

We estimate the 80% uncertainty range in these reserve estimates as being $\pm 10\%$ of the WEC⁴² estimates or the range of estimates in the literature^{4,43–46}, whichever is greater, for individual classes of reserves. We combine these reserve uncertainties with the provided 95% ranges of calorific values and emission factors for each class of energy reserves⁶ (Supplementary Table 3). See Supplementary Information for an expanded description of the methods.

31. Levitus, S., Antonov, J. & Boyer, T. Warming of the world ocean, 1955–2003. *Geophys. Res. Lett.* **32**, L02604, doi:10.1029/2004GL021592 (2005).
32. Knutti, R. & Tomassini, L. Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophys. Res. Lett.* **35**, L09701, doi:10.1029/2007GL032904 (2008).
33. Knutti, R., Stocker, T. F., Joos, F. & Plattner, G. K. Probabilistic climate change projections using neural networks. *Clim. Dyn.* **21**, 257–272 (2003).
34. Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A. & Rayner, N. A. An observationally based estimate of the climate sensitivity. *J. Clim.* **15**, 3117–3121 (2002).
35. Forest, C. E., Stone, P. H. & Sokolov, A. P. Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett.* **33**, L01705, doi:10.1029/2005GL023977 (2006).
36. Andronova, N. G. & Schlesinger, M. E. Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.* **106**, D19 22605–22611 (2001).
37. Piani, C., Frame, D. J., Stainforth, D. A. & Allen, M. R. Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophys. Res. Lett.* **32**, L23825, doi:10.1029/2005GL024452 (2005).
38. Murphy, J. M. et al. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772 (2004).
39. Annan, J. D. & Hargreaves, J. C. Using multiple observationally-based constraints to estimate climate sensitivity. *Geophys. Res. Lett.* **33**, L06704, doi:10.1029/2005GL025259 (2006).
40. Hegerl, G. C., Crowley, T. J., Hyde, W. T. & Frame, D. J. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* **440**, 1029–1032 (2006).
41. Knutti, R., Meehl, G. A., Allen, M. R. & Stainforth, D. A. Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Clim.* **19**, 4224–4233 (2006).
42. Clarke, A. W. & Trinnaman, J. A. (eds) *Survey of Energy Resources 2007* (World Energy Council, 2007).
43. BP. *BP Statistical Review of World Energy June 2008* (BP, London, 2008); available at (www.bp.com/statisticalreview).
44. Rempe, H., Schmidt, S. & Schwarz-Schampera, U. *Reserves, Resources and Availability of Energy Resources 2006* (German Federal Institute for Geosciences and Natural Resources, 2007).
45. Abraham, K. International outlook: world trends: Operators ride the crest of the global wave. *World Oil* **228**, no. 9 (2007).
46. Radler, M. Special report: Oil production, reserves increase slightly in 2006. *Oil Gas J.* **104**, 20–23 (2006); available at (<http://www.ogj.com/currentissue/index.cfm?p=7&v=104&i=47>).

Warming caused by cumulative carbon emissions towards the trillionth tonne

Myles R. Allen¹, David J. Frame^{1,2}, Chris Huntingford³, Chris D. Jones⁴, Jason A. Lowe⁵, Malte Meinshausen⁶ & Nicolai Meinshausen⁷

Global efforts to mitigate climate change are guided by projections of future temperatures¹. But the eventual equilibrium global mean temperature associated with a given stabilization level of atmospheric greenhouse gas concentrations remains uncertain^{1–3}, complicating the setting of stabilization targets to avoid potentially dangerous levels of global warming^{4–8}. Similar problems apply to the carbon cycle: observations currently provide only a weak constraint on the response to future emissions^{9–11}. Here we use ensemble simulations of simple climate-carbon-cycle models constrained by observations and projections from more comprehensive models to simulate the temperature response to a broad range of carbon dioxide emission pathways. We find that the peak warming caused by a given cumulative carbon dioxide emission is better constrained than the warming response to a stabilization scenario. Furthermore, the relationship between cumulative emissions and peak warming is remarkably insensitive to the emission pathway (timing of emissions or peak emission rate). Hence policy targets based on limiting cumulative emissions of carbon dioxide are likely to be more robust to scientific uncertainty than emission-rate or concentration targets. Total anthropogenic

emissions of one trillion tonnes of carbon (3.67 trillion tonnes of CO₂), about half of which has already been emitted since industrialization began, results in a most likely peak carbon-dioxide-induced warming of 2 °C above pre-industrial temperatures, with a 5–95% confidence interval of 1.3–3.9 °C.

Under conventional climate stabilization scenarios, greenhouse gas emissions are reduced until atmospheric composition approaches a stabilization level consistent with a desired equilibrium warming and are then adjusted to hold concentrations stable thereafter⁵. If climate system and carbon cycle properties were known, this would be straightforward: we could reliably map emissions to temperatures and vice versa. For example, if the climate system were to follow the response of a simple model with most likely values of key parameters (see Methods Summary and Supplementary Information), the emissions scenario highlighted by the solid red line in Fig. 1a would bring atmospheric carbon dioxide (CO₂) concentrations towards 490 p.p.m. (parts per million) by the end of the twenty-first century (solid red line in Fig. 1b). Under the ‘490 p.p.m. stabilization scenario’ shown by the dotted red lines, rapid reductions cease after 2070, with smaller subsequent adjustments causing concentrations to converge to 490 p.p.m.

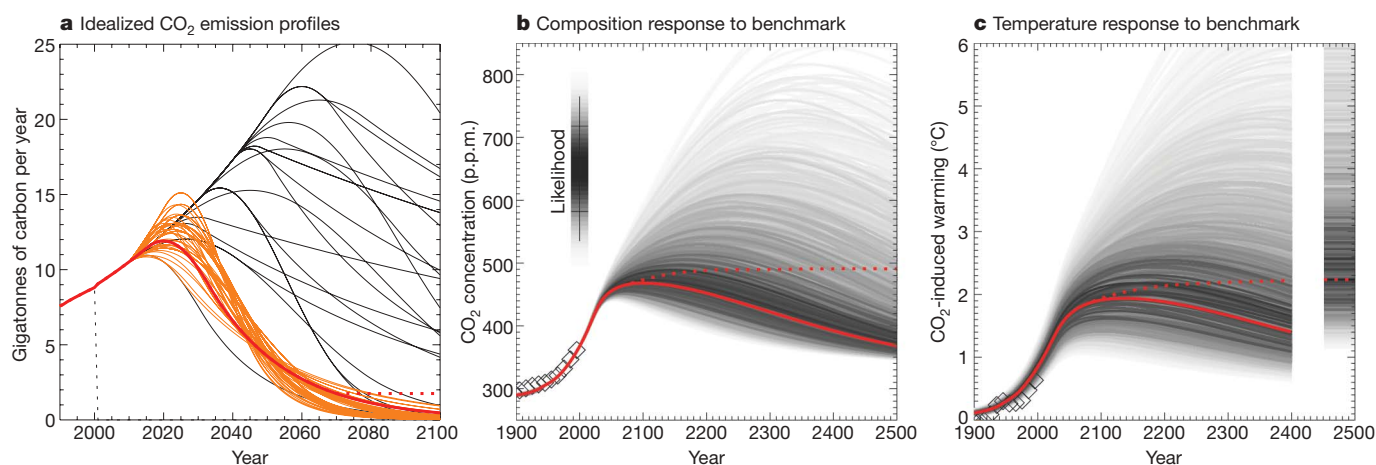


Figure 1 | Idealized carbon dioxide emission scenarios and response to benchmark scenario. **a**, Emissions, including zero emissions after 2000 (dotted black line). Solid red and orange lines show scenarios with cumulative emissions 1750–2500 within 1% of 1 Tt C. Solid red line shows benchmark case and dotted red line shows the ‘490 p.p.m. stabilization’ scenario. **b**, CO₂ concentration response to benchmark scenario with best-fit combination of simple climate model parameters (solid red line) and with random parameter combinations shaded by likelihood (grey plume). The vertical scale bar shows the corresponding likelihood profile for a normally

distributed quantity, with black line showing 5–95% (horizontal tickmarks: 17–83%) confidence interval. The dotted red line shows best-fit response to stabilization scenario. **c**, Temperature response to benchmark scenario from simple model: best fit in red and likelihood profile in grey. Bar on right shows likelihood profile for peak warming response to ‘490 p.p.m. stabilization’ emissions scenario: in cases where temperatures are still rising in 2500, equilibrium warming response to 2500 CO₂ concentration is plotted. Diamonds in **b** and **c** show observed CO₂ concentrations and temperatures (relative to 1900–1920), respectively.

¹Department of Physics, University of Oxford, OX1 3PU, UK. ²Smith School of Enterprise and the Environment, University of Oxford, OX1 2BQ, UK. ³Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK. ⁴Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK. ⁵Met Office Hadley Centre (Reading Unit), Department of Meteorology, University of Reading, RG6 6BB, Reading, UK. ⁶Potsdam Institute for Climate Impact Research, 14412 Potsdam, Germany. ⁷Department of Statistics, University of Oxford, OX1 3TG, UK.

over the following century⁵. With most likely values of key parameters in this model (including an Equilibrium Climate Sensitivity, or ECS, of 2.8 °C for doubling atmospheric CO₂) these emissions cause a warming of 2.2 °C above pre-industrial by 2500, but much higher responses (shown by the shaded bar on the right of Fig. 1c) are also consistent with current uncertainties in ECS^{1,2} and carbon cycle^{9–11}. The shading shows the range arising solely from known uncertainties in current feedbacks: the true uncertainty is greater, particularly for long time-scales and higher responses, because feedbacks may change as the climate changes³. In practice, our descendants would be unlikely to adhere to this specific emissions scenario in the event that both CO₂ and temperature overshoot the original targets, but this illustrates the 'stabilization dilemma': either we specify a temperature or concentration target and accept substantial uncertainty in the emissions required to achieve it or we specify emissions and accept even more uncertainty in the temperature response.

There is, however, no intrinsic reason why emissions must be adjusted to maintain a target concentration, particularly if temperatures then look likely to overshoot. Instead, we could assume that once the current rate of increase in emissions starts to fall, it declines until it reaches a maximum annual percentage rate of decrease which is subsequently maintained indefinitely (see Methods and Supplementary Fig. S3a). Varying both the timing of the transition and the maximum rate of decrease gives idealized emission profiles (solid lines in Fig. 1a) which can reproduce the essential features of most mitigation scenarios over the next few decades. Unlike conventional stabilization and overshoot scenarios^{5,12,13}, the integrals under these 'containment scenarios', or cumulative total carbon dioxide emissions over the entire 'anthropocene' period, are bounded. For integrals less than two trillion tonnes of carbon (TtC) almost all emissions occur before 2200. The solid red line in Fig. 1a is one such containment scenario with an integral of 1 TtC: with most likely values of climate system properties applied to the simple climate model, it suggests that both carbon dioxide levels and temperatures peak (at 470 p.p.m. and almost exactly 2 °C above pre-industrial respectively) and then decline thereafter (solid red lines in Fig. 1b and c). Substantially higher responses to this benchmark scenario are also consistent with current observations, as shown by the spread of the grey shaded plumes, but even the most pessimistic show temperatures declining within a couple of centuries, unlike the response to the dotted 'stabilization scenario' emissions, under which CO₂

levels and temperatures could continue rising for centuries if the climate turns out to be more sensitive than our current best estimate.

The maximum warming under concentration overshoot scenarios^{12,13} is better constrained by observations than the long-term response to a stabilization scenario¹⁴, but such overshoot scenarios appear to require at least three specified targets: the final concentration, the size of the overshoot and its duration. A simpler target is suggested by Fig. 2, which plots peak carbon-dioxide-induced warming against the total carbon dioxide released over the entire period 1750 to 2500, expressed as TtC, for 250 containment scenarios (a subset of which are shown by the solid black lines in Fig. 1a). Each white cross in Fig. 2 shows maximum warming under one scenario in the simple climate model with most likely values of model parameters. The black–grey–white shading denotes the relative likelihood of different levels of warming for the same total carbon dioxide released, allowing for uncertainty in modelled carbon cycle, atmosphere and ocean. The crosses all lie close to a single curve despite the fact that peak emission rates differ by up to a factor of two for the same cumulative carbon release. The total emissions determine peak CO₂-induced warming under a containment scenario, not the peak emission rate or other details of the emission pathway. We focus here on peak warming, but we find this result also applies to average warming over the 2000–2500 period or indeed warming by any given date after emissions have substantially declined (see Supplementary Information).

This conclusion is supported by an independent model and approach. The coloured diamonds in Fig. 2 show peak warming as a function of cumulative CO₂ emissions as simulated by different versions of the Hadley Centre Simple Climate-Carbon-Cycle Model (HadSCCCM1)¹⁰ for a subset of these 250 containment scenarios. Different colours show the impact of adjusting HadSCCCM1 parameters to fit the behaviour of the eleven coupled Earth System Models (ESMs)¹⁵ in the Coupled Climate Carbon Cycle Model Intercomparison (C⁴MIP) experiment¹⁶. The fit becomes less reliable for temperatures more than 0.5 °C above the range simulated by the corresponding C⁴MIP simulation: hence diamonds are plotted only where peak warming is inside this range. Again, symbols corresponding to each particular ESM fall close to a single line: peak warming is independent of the shape of the emissions path and depends only on the cumulative total.

This insensitivity to the timing of future emissions suggests we can define the Cumulative Warming Commitment (CWC) as the peak

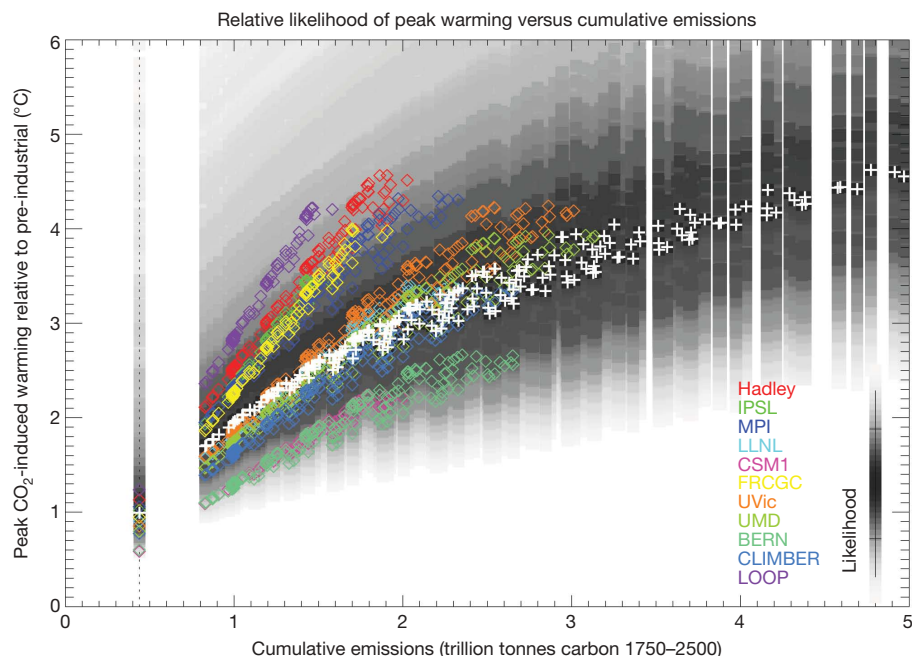


Figure 2 | Peak CO₂-induced warming as a function of total cumulative emissions 1750–2500 for 250 idealized emission scenarios. (A subset is plotted as black lines in Fig. 1a.) White crosses correspond to best-fit values of simple climate model parameters, with each cross corresponding to a single scenario. Grey shading shows relative likelihood of other parameter combinations, plotted in order of increasing likelihood, showing the uncertainty in peak warming arising from parameter uncertainty in the simple model. Coloured diamonds show responses of the HadSCCCM1 model with parameters fitted to ESMs in the C⁴MIP experiment, with colours indicating the corresponding ESM. Diamonds are plotted only where temperatures remain within 0.5 °C of the range of the tuning data set (the SRES A2 scenario) to ensure a valid emulation. Bar and symbols at 4.4 TtC show peak warming assuming zero emissions after 2000. Likelihood scale bar as in Fig. 1b.

warming response to a given total injection of CO₂ into the atmosphere following our best estimate of anthropogenic emissions to date and any future emissions pathway that is smooth, positive and ends in exponential decline. CWC provides a simple measure of climate system response to scenarios in which CO₂ concentrations peak and decline. Unlike ECS, CWC relates emissions right through to temperature, so a range on CWC also incorporates uncertainty in the carbon cycle.

The coloured diamonds in Fig. 2 suggest that the ratio of CWC per trillion tonnes of carbon emitted, or normalized CWC, is approximately constant across the range of responses over which we can calibrate HadSCCCM1 against the C⁴MIP ensemble (the lines are nearly straight), although it does depend on which member of the C⁴MIP ensemble we emulate. The white crosses generated by the simple model suggest that this ratio declines slightly for cumulative emissions exceeding 2 Tt C. For the highest scenarios, emissions continue beyond 2500 and all results for high emission totals should be treated with caution: warmer temperatures increase the likelihood of strongly nonlinear feedbacks in the climate system¹⁷ that might be represented only in more comprehensive models¹⁸.

Figure 3 shows an analysis of uncertainty in CWC for the specific case of 1 Tt C total anthropogenic emission of CO₂. The 49 solid red and orange curves (superimposed and almost indistinguishable) show 'likelihood profiles'¹⁹ for peak warming response to the red and orange containment scenarios plotted in Fig. 1a, all of which represent cumulative emissions over 1750–2500 that fall within 1% of 1 Tt C. These show the relative likelihood of the most likely versions of the simple climate model out of the subset that gives the values of peak warming shown on the horizontal axis, where likelihoods are computed from the constraints detailed in Supplementary Figs S1 and S2. The overall best-fit (most likely) value of the 1 Tt C CWC is 2 °C, and the 5–95% confidence interval (given by the range over which likelihoods exceed the corresponding threshold¹⁹) is 1.3–3.9 °C, again independent of which scenario is used to estimate it. The black dotted curve shows the likelihood profile for peak warming in response to the '490 p.p.m. stabilization scenario': the higher emissions in this scenario after 2070 have little impact on the most likely peak warming (they prolong the peak rather than raising it), but they double the likelihood of warming in the 3–6 °C range because CO₂ levels and temperatures continue to rise in more sensitive versions of the model.

Coloured symbols in Fig. 3 show CWC under these 49 1 Tt C containment scenarios predicted by HadSCCCM1 emulating the C⁴MIP ESMs. The horizontal spread of each set of coloured symbols is small, reiterating that CWC does not depend on the shape of the emission pathway. The variation in emission pathway is illustrated by the vertical position of coloured symbols, showing the fractional reduction in emissions from 2010 to 2050 (right axis) for each of the 49 scenarios. The large vertical spread shows that very different emission pathways with the same cumulative total give the same peak warming: reductions by 2050 only matter insofar as they affect the total CO₂ released.

The black horizontal error bar in Fig. 3 shows a 5–95% Bayesian posterior probability interval for 1 Tt C CWC estimated from an additional independent model and approach detailed in ref. 20 using their representative distribution for climate system properties and our benchmark scenario. Corresponding intervals for the other 48 1 Tt C containment scenarios are almost identical, providing further evidence that the timing of emissions has no impact on CWC. The lower bound is consistent with the other two approaches detailed here. The upper bound is lower primarily because the posterior upper bound on past CO₂-attributable warming implied by ref. 20, although consistent with typical inter-model ranges for Transient Climate Response¹, is lower than the upper bound used to constrain the simple model, which is more consistent with observationally constrained confidence intervals for the transient response^{1,14,21}.

The need to limit the total amount of carbon dioxide released into the atmosphere has been noted before^{4,22–24}: the concept of CWC

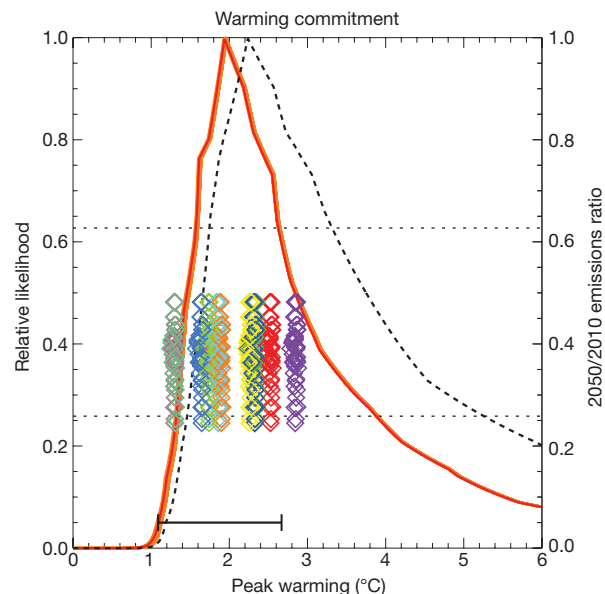


Figure 3 | Warming commitment for selected scenarios shown in Fig. 1a. Red and orange curves (superimposed) show likelihood profiles for cumulative warming commitment (CWC) under 49 scenarios with total emissions within 1% of 1 Tt C estimated with the simple climate-carbon-cycle model; dotted black curve shows likelihood profile for peak warming under '490 p.p.m. stabilization' scenario. Horizontal dotted lines show thresholds for the 17–83% and 5–95% confidence intervals¹⁹. The horizontal location of coloured symbols shows corresponding 1 Tt C CWCs predicted by HadSCCCM1 fitted to the ESMs in the C⁴MIP ensemble. The vertical location and right-hand axis shows fractional decrease in emissions 2010 to 2050 for each of the 49 emissions scenarios. Solid error bar shows 5–95% Bayesian posterior probability interval (1.1–2.7 °C) for 1 Tt C CWC based on an independent model and approach²⁰.

could provide the scientific basis for such limits. Because of its emissions path independence, CWC is a relatively simple quantity for policy discussions²⁵. It is also easier to constrain with observations than either the ECS or the emissions required to achieve a given stabilization target. The CWC is closely related to the response to a pulse injection used to define the Global Temperature Potential (GTP)^{26,27} and to the Zero Emissions Commitment (ZEC)^{18,28}, or additional warming that occurs after a sudden and complete cessation of emissions. GTP is defined for a particular timescale and background scenario, so CWC may be thought of as a 'peak GTP' averaging over scenarios. If we assume CWC is linear in cumulative emissions, then the sum of the CO₂-attributable warming in 2000 (0.85 °C with a 5–95% range of 0.6–1.1 °C) and the ZEC in 2000, both divided by best-guess cumulative emissions to 2000 (0.44 Tt C), could provide an estimate of normalized CWC. If we could further assume that the ZEC were negligible, then the first term alone, which corresponds to the 'Carbon-Climate Response' of ref. 25 and is independent of our simple models, implies a best-guess normalized CWC of 1.9 °C per TtC with a 5–95% range of 1.4–2.5 °C per TtC.

Hence, the asymmetry in our full range for CWC (Fig. 3) primarily arises from the possibility of a substantial ZEC. This emerges in our simple models because both CO₂ and temperatures take many decades to fall after emissions cease even for values of ECS around 3 °C, consistent with results from some more complex models^{4,18,28,29}. With the longer response times associated with higher, but still not particularly unlikely, values of ECS (for example, 4.5 °C), temperatures can rise substantially even after emissions are set to zero. Hence we argue that CWC allowing for a non-zero ZEC provides a more conservative basis for policy than the CWC neglecting ZEC, as proposed by ref. 25, unless further research rules out a substantial ZEC (in which case the upper bound on CWC would fall considerably). Reducing uncertainty in past warming attributable to greenhouse gases would also reduce

uncertainty in CWC^{21,30}, but similar-magnitude reductions in uncertainty in ocean heat uptake³⁰ and long-term carbon cycle properties have less impact (see Supplementary Table and Supplementary Discussion). We can therefore expect uncertainty in CWC to decline as the attributable warming signal strengthens (results here only use data to 2000).

CO₂-induced warming only equals total anthropogenic warming if the net effect of non-CO₂ anthropogenic climate forcing agents is relatively small. This may be the case at present, but it is unlikely to remain so. Another study²⁰ argues that non-CO₂ forcing reduces the budget of CO₂ emissions consistent with a best-guess total anthropogenic warming of 2 °C to just under 0.4 TtC for the 2000–2050 period, compared with a budget for this period of 0.4–0.5 TtC in the scenarios that we find give a most likely warming of 2 °C due to CO₂ alone. The balance of warming and cooling by non-CO₂ climate drivers is, however, strongly scenario-dependent. Defining a timescale- and scenario-independent method of combining the impact of short- and long-lived climate forcing agents is impossible²⁷ because emission rates determine the impact of agents with atmospheric lifetimes shorter than a few decades (the thermal response time of the climate system), whereas cumulative emissions drive the impact of CO₂. CO₂ emission rates, within the same cumulative total, have little impact on projected warming. A simpler policy framework might therefore be to limit emission rates of shorter-lived agents to avoid dangerous rates of warming and to use the concept of CWC to limit cumulative emissions of CO₂ (and other very-long-lived agents) to avoid a dangerous total warming commitment.

METHODS SUMMARY

We generate idealized carbon dioxide emission scenarios by continuously varying the fractional rate of change in emissions to give a smooth transition from exponential growth to exponential decline: see Supplementary Fig. S3a. With the exception of the coloured symbols in Figs 2 and 3, all results are based on a simple coupled climate carbon-cycle model detailed in the Supplementary Information with five free parameters: climate sensitivity; ocean thermal diffusivity; ocean/biosphere carbon uptake diffusivity; rate of advection of carbon into the deep ocean; and feedback parameter for carbon released by surface warming. These are subject to five constraints: warming attributable to greenhouse gases over the twentieth century²¹; effective ocean/troposphere heat capacity^{14,30}; observed net airborne fraction over the 1960–2000 period³¹; contribution of temperature–carbon-cycle feedbacks to net airborne fraction 1750–2100, based on the C⁴MIP ensemble¹⁶; and rate of uptake of carbon into the deep ocean based on current ESMs²⁹. Parameters are varied at random and the likelihood of each parameter combination is evaluated as the product of the likelihoods with respect to the individual constraints normalized to unity at the best-fit value. Relative likelihoods can then be plotted against any variable simulated by the model, and the outline, or likelihood profile¹⁹, evaluated against standard thresholds to give confidence intervals. The grey-shaded points in Figs 1 and 2 are plotted in order of increasing likelihood to visualize the evolution of the likelihood profile as a function of time or total CO₂ emissions. The coloured symbols in Figs 2 and 3 represent 11 versions of the HadSCCM1 model^{10,15} fitted to the ESMs in the C⁴MIP study¹⁶, with the ESM name indicated by the legend in Fig. 2. The spread illustrates the range of behaviour of current ESMs but it is not a comprehensive measure of model uncertainty.

Received 25 September 2008; accepted 25 March 2009.

1. Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) Ch. 10 (Cambridge Univ. Press, 2007).
2. Knutti, R. & Hegerl, G. C. The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nature Geosci.* **1**, 735–743 (2008).
3. Hansen, J. E. *et al.* Target atmospheric CO₂: Where should humanity aim? *Open Atmos. Sci. J.* **2**, 217–231 (2009).
4. Solomon, S., Plattner, G.-K., Knutti, R. & Friedlingstein, P. Irreversible climate change due to carbon dioxide emissions. *Proc. Natl Acad. Sci. USA*. DOI:10.1073/pnas.0812721106 (2009).
5. Wigley, T. M. L., Richels, R. & Edmonds, J. A. Economic and environmental choices in the stabilisation of atmospheric CO₂ concentrations. *Nature* **379**, 240–243 (1996).

6. Schneider, S. H. & Mastrandrea, M. D. Probabilistic assessment of dangerous climate and emissions pathways. *Proc. Natl Acad. Sci. USA* **102**, 15728–15735 (2005).
7. Meinshausen, M. in *Avoiding Dangerous Climate Change* (eds Schellnhuber, H. J. *et al.*) Ch. 28 (Cambridge Univ. Press, 2006).
8. Harvey, L. D. D. Allowable CO₂ concentrations under the United Nations Framework Convention on Climate Change as a function of the climate sensitivity probability distribution function. *Environ. Res. Lett.* **2**, 014001 (2007).
9. Friedlingstein, P. *et al.* How positive is the feedback between climate change and the carbon cycle? *Tellus* **55B**, 692–700 (2003).
10. Jones, C. D., Cox, P. M. & Huntingford, C. Climate-carbon cycle feedbacks under stabilization. *Tellus* **58B**, 603–613 (2006).
11. Matthews, H. D. Effect of CO₂ fertilization uncertainty on future climate change in a coupled climate-carbon model. *Glob. Change Biol.* **13**, 1068–1078 (2007).
12. Den Elzen, M. G. J. & Van Vuuren, D. P. Peaking profiles for achieving long-term temperature targets with more likelihood at lower costs. *Proc. Natl Acad. Sci. USA* **104**, 17931–17936 (2007).
13. Huntingford, C. & Lowe, J. Overshoot scenarios and climate change. *Science* **316**, 829 (2007).
14. Frame, D. J. *et al.* Alternatives to stabilization scenarios. *Geophys. Res. Lett.* **33**, DOI:10.1029/2006GL025801 (2006).
15. Huntingford, C. *et al.* Contributions of thermal and carbon cycle uncertainty to future climate projection spread. *Tellus* **61B**, 355–360 (2009).
16. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis, results from the C4MIP model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
17. Schneider, S. H. Abrupt non-linear climate change, irreversibility and surprise. *Glob. Environ. Change* **14**, 245–258 (2004).
18. Lowe, J. A. *et al.* How difficult is it to recover from dangerous levels of global warming? *Environ. Res. Lett.* **4**, 014012 (2009).
19. Pawitan, Y. In *all Likelihood: Statistical Modeling and Inference Using Likelihood* Ch. 2.6 and 3.4 (Oxford Univ. Press, 2001).
20. Meinshausen, M. *et al.* Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature* doi:10.1038/nature08017 (this issue).
21. Stott, P. A. *et al.* Observational constraints on past attributable warming and predictions of future global warming. *J. Clim.* **19**, 3055–3069 (2006).
22. Broecker, W. S. CO₂ arithmetic. *Science* **315**, 1371 (2007).
23. Wigley, T. M. L. CO₂ emissions: a piece of the pie. *Science* **316**, 829–830 (2007).
24. Matthews, H. D. & Caldeira, K. Stabilizing climate requires near-zero emissions. *Geophys. Res. Lett.* **35**, L04705 (2009).
25. Matthews, H. D. *et al.* The proportionality of global warming to cumulative carbon emissions. *Nature* doi:10.1038/nature08047 (in the press).
26. Shine, K. P. *et al.* Alternatives to the global warming potential for comparing climate impacts of emissions of greenhouse gases. *Clim. Change* **68**, 281–302 (2005).
27. Shine, K. P. *et al.* Comparing the climate effect of emissions of short- and long-lived climate agents. *Phil. Trans. R. Soc. A* **365**, 1903–1914 (2007).
28. Friedlingstein, P. & Solomon, S. Contributions of past and present human generations to committed warming caused by carbon dioxide. *Proc. Natl Acad. Sci. USA* **102**, 10832–10836 (2005).
29. Plattner, G.-K. *et al.* Long-term climate commitments projected with climate-carbon cycle models. *J. Clim.* **21**, 2721–2751 (2008).
30. Frame, D. J. *et al.* Constraining climate forecasts: the role of prior assumptions. *Geophys. Res. Lett.* **32**, DOI:10.1029/2004GL022241 (2005).
31. Keeling, C. D. & Whorf, T. P. Atmospheric CO₂ records from sites in the SIO air sampling network. In *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US DOE, 2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Gillett, K. Shine and T. Stocker for suggestions, P. Stott for estimates of twentieth-century-attributable warming, J. Welby for help calibrating the simple climate model, P. Friedlingstein and the C⁴MIP modelling community for model output and I. Tracey for help with the manuscript. M.R.A. and D.J.F. acknowledge support from NERC and the FP6 ENSEMBLES project. M.R.A. received additional support from the International Detection and Attribution Working Group (IDAG), supported by the DOE Office of Science, Office of Biological and Environmental Research and NOAA Climate Program Office, and from the British Council. C.H. acknowledges the CEH Science Budget Fund. C.D.J. and J.A.L. were supported by the Joint DECC, Defra and MoD Integrated Climate Programme (DECC/Defra GA01101; MoD CBC/2B/0417_Annex C5).

Author Contributions M.R.A. and D.J.F. designed, tested and ran the simple climate model. C.H., C.D.J. and J.A.L. developed and tuned HadSCCM1 and C.H. and J.A.L. ran the simulations; M.M. ran the MAGICC model contributing to Fig. 3 and N.M. advised on statistical analysis. All authors contributed to writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.R.A. (myles.allen@physics.ox.ac.uk).

Evolutionary diversification in stickleback affects ecosystem functioning

Luke J. Harmon^{1,2*}, Blake Matthews^{2,3*}, Simone Des Roches¹, Jonathan M. Chase⁴, Jonathan B. Shurin² & Dolph Schluter²

Explaining the ecological causes of evolutionary diversification is a major focus of biology, but surprisingly little has been said about the effects of evolutionary diversification on ecosystems^{1–3}. The number of species in an ecosystem and their traits are key predictors of many ecosystem-level processes, such as rates of productivity, biomass sequestration and decomposition^{4,5}. Here we demonstrate short-term ecosystem-level effects of adaptive radiation in the threespine stickleback (*Gasterosteus aculeatus*) over the past 10,000 years. These fish have undergone recent parallel diversification in several lakes in coastal British Columbia, resulting in the formation of two specialized species (benthic and limnetic) from a generalist ancestor⁶. Using a mesocosm experiment, we demonstrate that this diversification has strong effects on ecosystems, affecting prey community structure, total primary production, and the nature of dissolved organic materials that regulate the spectral properties of light transmission in the system. However, these ecosystem effects do not simply increase in their relative strength with increasing specialization and species richness; instead, they reflect the complex and indirect consequences of ecosystem engineering by sticklebacks. It is well known that ecological factors influence adaptive radiation^{7,8}. We demonstrate that adaptive radiation, even over short timescales, can have profound effects on ecosystems.

Most previous studies of adaptive radiation have focused on the effects of environmental factors on diversification. For example, decades of work in 'model' systems of adaptive radiation, like Caribbean *Anolis* lizards⁹, Galapagos finches¹⁰ and African cichlids⁸, have revealed how species' environments shape speciation and adaptation. Ecological theory suggests that diversification of one into many species, each with distinct environmental roles, can affect prey diversity, energy flow, food chain length and other aspects of ecosystems^{5,7,8,11,12}. Adaptive diversification leading to greater predator diversity, for example, could increase top-down control over a wider range of consumer species and strengthen cascading effects of predators on lower trophic levels^{13,14}. It follows that adaptive radiation may modify the environmental conditions of ecosystems and shape the selective pressures of other species^{12,15}. However, in general, the consequences of evolutionary diversification for ecosystems are largely unknown.

We introduced threespine stickleback into outdoor mesocosms to study the short-term effect of speciation and specialization of stickleback on pond ecosystems. These fish are common in marine coastal waters of the Northern Hemisphere, and have colonized and adapted to freshwater habitats in coastal areas¹⁶. In seven lakes of British Columbia, sympatric species pairs occupy different niches within the same lake: one limnetic and the other benthic. These two forms

have evolved independently multiple times in the region over the past 10,000 years¹⁷, they differ by morphology and habitat use, and they are reproductively isolated with low levels of gene flow¹⁸. Most relevant to their effects on ecosystems, the two forms have distinct diets: limnetic and benthic forms specialize on pelagic zooplankton and littoral invertebrates, respectively. In contrast, sticklebacks that have not diversified, and occur alone in similar small lakes, have an intermediate form and a more generalist diet^{6,7} (see Supplementary Fig. 1). Previous studies have found that sticklebacks can cause trophic cascades in pond ecosystems¹⁹. Here we address whether the evolutionary diversification of stickleback populations affects ecosystem function.

We conducted an experiment to test the effects of evolutionary diversification on ecosystems using benthic and limnetic fish from Paxton Lake—one of the seven lakes with a sympatric species pair¹⁷—and an intermediate generalist fish from a nearby single-species lake, Cranby Lake. We introduced the fish into mesocosms (1,136 l) according to four treatments, each replicated ten times: generalist (G; Cranby Lake fish), limnetic (L), benthic (B), and species pair (LB; all from Paxton Lake). Treatments G and BL represent combinations that occur in natural lakes, whereas the B and L treatments were used to evaluate the effects of specialization independent from an increase in species richness. Each mesocosm initially contained the same total biomass of fish (between 5.0 and 6.0 g); because stickleback forms differ in body size, each tank contained between three and six fish (G = 4, B = 3, L = 6, BL = 5). Mesocosms contained both benthic and pelagic zones and were seeded with sediments and invertebrates from nearby ponds inhabited by sticklebacks. Because predators can influence ecosystems through both trophic and non-trophic interactions^{20,21}, we examined whether stickleback diversification could influence a wide variety of ecosystem properties and functions over a 10-week period (see Supplementary Methods for details). We evaluated differences in these measures among treatments, with a focus on overall effects and two sets of planned contrasts: between G and BL, to test the overall effects of evolutionary diversification, and between BL and either B or L, to test for the effects of diversification versus specialization alone.

We tested two possible hypotheses for the consequences of stickleback diversification on aquatic food webs and ecosystems (see Supplementary Fig. 1). Our first hypothesis, arising from trophic cascade theory (H1), was that sticklebacks would cause the strongest cascades in their preferred foraging habitats, leading to higher levels of primary productivity with increasing predator specialization (G < B, L < BL). Our second hypothesis was that distinct types of sticklebacks might differ in their engineering effects on ecosystems (H2).

¹Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844-3051, USA. ²Biodiversity Research Centre and Zoology Department, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ³EA-WAG, Aquatic Ecology, Seestrasse 79, 6047 Kastanienbaum, Switzerland. ⁴Department of Biology, Washington University, St Louis, Missouri 63130, USA.

*These authors contributed equally to this work.

Sticklebacks are part of a complex food web and could affect ecosystems by a whole suite of trophic and non-trophic interactions, including indirect mechanisms such as habitat modification and nutrient recycling²². In particular, on the basis of previous research on ecosystem engineering by fish^{23,24}, we hypothesized that sticklebacks could affect light transmission in the water column by affecting algal biomass through a combination of cascading trophic interactions and non-trophic engineering interactions (for example, nutrient recycling, habitat modification). Fish can affect light transmission in the water column by indirectly influencing the abundance of suspended algal particles which both directly block light²⁴ and produce dissolved organic substances that attenuate light²⁵ (see Supplementary Fig. 1). Dissolved organic carbon (DOC) produced by algae, compared to that released from the decomposition of leaf litter and sediment organic matter, is more easily degradable by bacteria and more transparent to light in the ultraviolet range²⁶. As such, our second hypothesis was that sticklebacks would have indirect effects on algae, DOC and the light transmission properties of the water; however, we made no specific predictions about whether diversification or specialization of fish would have larger ecosystem-engineering effects.

We found that stickleback diversification affected the community composition of organisms at lower trophic levels. We found strong effects of sticklebacks on pelagic prey communities (Fig. 1a; multivariate analysis of variance (MANOVA) on two-dimensional non-metric multidimensional scaling (NMDS) using zooplankton species' total biomass per tank, Wilks' $\lambda = 0.65$, $P = 0.02$), with a significant contrast between treatments representing the effects of stickleback diversification (G versus BL, Hotelling $T^2 = 10.9$, $P = 0.02$). Contrasts comparing the effects of specialization versus diversification attribute this effect to specialization alone and, specifically, to the transition from the generalist to the limnetic form. Pelagic prey communities with two specialists were significantly different compared with those with only the benthic form (B versus BL, NMDS Euclidean distance 0.60, $T^2 = 9.9$, $P = 0.02$) but not from those with only the limnetic form (L versus BL, NMDS Euclidean distance 0.17, $T^2 = 0.7$, $P = 0.7$). The calanoid copepod (*Skistodiaptomus*

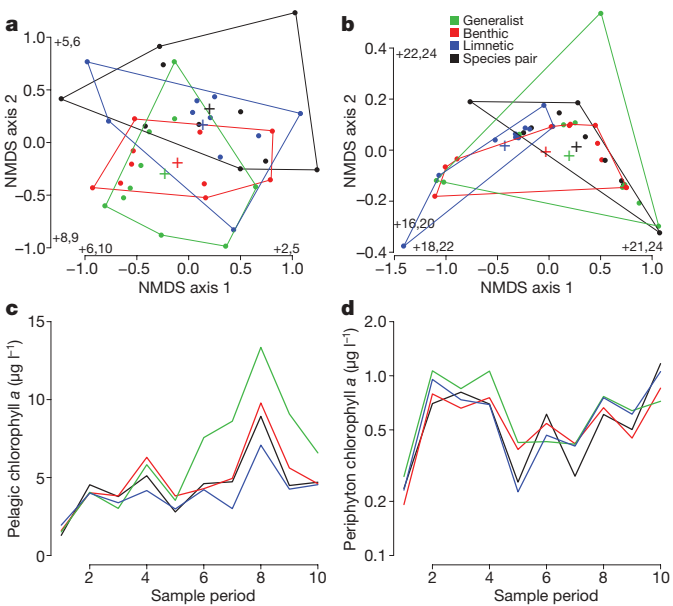


Figure 1 | Differences in community structure and productivity among treatments. **a, b**, First two non-metric multidimensional scaling (NMDS) axes of pelagic (**a**) and benthic (**b**) invertebrate biomass. Points represent individual mesocosms, coloured by treatments, crosses represent treatment means, and polygons surround all points for a given treatment. Numbers on each axis correspond to the two species with the strongest positive and negative loadings (see Table 1). **c, d**, Amount of chlorophyll in pelagic (phytoplankton, **c**) and benthic (periphyton, **d**) zones at ten sampling periods through the course of the experiment. Coloured lines represent treatment means weekly throughout the experiment.

oregonensis) was virtually eliminated from tanks with limnetics (that is, both the L and LB treatments, Table 1). In natural lakes, this zooplankton species is highly transparent and vertically migrates through the water column to avoid stickleback predation, but, nevertheless, it is a common diet item for both limnetics and pelagic specialists²⁷. Hence, predator specialization over the course of an

Table 1 | Species, average lengths and average densities (per litre) for pelagic and benthic organisms at the end of the experiment

	Size (mm)	Treatment			
		Generalist (G)	Benthic (B)	Limnetic (L)	Species pair (BL)
Pelagic					
(1) <i>Alonella</i>	0.25	678	625	337	862
(2) <i>Bosmina</i>	0.24	16	222	58	23
(3) <i>Chaoborus</i>	5.33	0	0	4	0
(4) Chironominae	0.86	33	67	49	14
(5) <i>Chydorus sphaericus</i>	0.24	120	119	144	444
(6) <i>Daphnia</i>	0.88	0	5	16	0
(7) <i>Diaphanosoma</i>	0.36	5	57	4	9
(8) <i>Sida crystallina</i>	0.53	0	5	0	0
(9) <i>Skistodiaptomus oregonensis</i>	0.73	552	351	53	14
(10) Cyclopoida	0.47	93	248	428	680
(11) Mite	0.37	0	0	4	0
(12) Nauplii	0.17	9,787	10,093	7,210	7,102
(13) Stone fly larva	0.98	0	26	0	0
(14) Surface insect	3.21	0	5	0	0
Benthic					
(15) Amphipoda	3.68	0	0	0	1
(16) Chironomidae	3.9	255	237	140	284
(17) Cladocera	1.1	2	7	4	26
(18) Coleoptera	2.76	0	1	1	0
(19) Diptera	3.35	0	0	0	1
(20) Gastropoda	2.02	2	9	10	5
(21) Hydracarina	0.26	0	1	1	1
(22) Odonata	7.7	2	0	0	1
(23) Ostracoda	0.52	0	104	10	28
(24) Pelecypoda	3.2	0	0	0	3
(25) Trichoptera pupae	5.35	0	0	7	0
(26) Other pupae	3.54	0	2	0	3

Table 2 | Results of profile analysis on measures of productivity and respiration

Test	Comparison	Test statistic	Phytoplankton	Periphyton	NPP	GPP	Respiration
Time	Overall	$F_{28,9}$	34.9§	99.9§	258.3§	228.7§	439.5§
	G versus BL	$F_{12,9}$	7.9†	19.2§	59.3§	51.8§	93.9§
	B versus BL	$F_{12,9}$	9.8§	20.1§	49.1§	46.2§	96.5§
	L versus BL	$F_{12,9}$	6.5†	21.2§	55.3§	50.0§	96.3§
Treatment	Overall	$F_{3,36}$	3.1†	0.3	4.4†	5.8†	6.3†
	G versus BL	$F_{1,18}$	3.8*	0.7	2.5	4.7†	8.0†
	B versus BL	$F_{1,18}$	0.1	0.1	3.2*	3.0*	1.9
	L versus BL	$F_{1,18}$	0.7	0.0	0.1	0	0.1
Time by treatment interaction	Overall	Wilks' λ	0.29*	0.39	0.31*	0.38	0.48
	G versus BL	T^2	38.9*	16.7	55.5†	21.2	24.9
	B versus BL	T^2	9.4	28.9	5.5	5.1	16.9
	L versus BL	T^2	12.6	22.2	22.2	24.1	10.7

The table entries represent values of test statistics. See Supplementary Information for details of the profile analysis.

* $P < 0.1$.

† $P < 0.05$.

‡ $P < 0.01$.

§ $P < 0.0001$.

adaptive radiation could potentially alter the strength of trophic interactions between prey and their resources. We found no differences among treatments in the composition of benthic invertebrate communities (Fig. 1b, MANOVA on two-dimensional NMDS using an index of total biomass (see Supplementary Information) for each species in each tank, Wilks' $\lambda = 0.8$, $P = 0.3$).

Treatments differed strongly in their gross primary productivity (GPP), net primary productivity (NPP) and respiration ($R = GPP - NPP$), with the strongest differences occurring between the G and BL treatments (Table 2). Treatments also differed in their levels of algal biomass (Table 2 and Fig. 1c), again with the strongest differences occurring between the G and BL treatments (Table 2). There were no differences detected between treatments in the amount of benthic algae (periphyton, Table 2 and Fig. 1d). Despite the strong effects of limnetics on pelagic zooplankton community structure, the treatments including this species (L and BL) did not have the highest algal biomass, contrary to our hypothesis based on trophic cascade theory (H1). Instead, the generalist treatment had the most algae (Fig. 1c), implying that neither specialization nor diversification led to stronger trophic cascades. Strong contrasts in the strength of cascades that affect algal biomass are more likely when zooplanktivorous fish vary in their foraging efficiency on predominantly herbivorous (for example, *Daphnia*) rather than omnivorous zooplankton taxa (for example, copepods). In the current experiment, *Daphnia* were reduced to very low abundance by sticklebacks in all treatments. As a result, the zooplankton communities differed primarily in their relative abundance of omnivorous and predation-resistant *Diaptomus* copepods (Table 1)²²; for this reason, we may not have detected a difference in the strength of trophic cascades among treatments.

To investigate the possible ecosystem engineering effects of sticklebacks we tested whether several ecosystem properties (see Supplementary Information) differed among stickleback treatments. We focused on the composition of DOC because it can profoundly affect the biology and physics of aquatic ecosystems by affecting the transmissibility of different wavelengths of light through the water^{25,26,28,29}. In agreement with the second hypothesis, we found that sticklebacks act as ecosystem engineers²³, strongly affecting the composition of the DOC pool and the physical light environment. Treatments were similar in their total amount of dissolved organic carbon (DOC, $F_{3,36} = 0.38$, $P = 0.8$; Fig. 2a), but varied in their DOC composition. By the end of the experiment, water differed significantly among treatments in its transmission of both photosynthetically available radiation (PAR: light with wavelengths between 400 and 700 nm; $F_{3,36} = 3.8$, $P = 0.02$; Fig. 2b) and ultraviolet radiation (A_{320} : $F_{3,36} = 6.3$, $P = 0.001$, Fig. 2c; this result was the same over the entire ultraviolet band (280–400 nm, see Supplementary Information)). Furthermore, dissolved substances fluoresced significantly

more at shorter excitation wavelengths in the generalist treatment (Lγ (low-molecular-mass molecules, <5,000 Da, that fluoresce at low excitation wavelengths); ANOVA of Lγ/total $F_{3,33} = 2.9$, $P = 0.05$, no planned contrasts were significant; Fig. 2d). As above, the planned contrasts revealed effects of both diversification and specialization on the light environment in the mesocosms (PAR: diversification, $t = -2.7$, $P = 0.02$, specialization, B versus BL, $t = -1.0$, $P = 0.3$, L versus BL, $t = -2.6$, $P = 0.02$; A_{320} : diversification, $t = -2.92$, $P = 0.009$, specialization, B versus BL, $t = -0.7$, $P = 0.5$, L versus BL, $t = -0.1$, $P = 0.9$). Differences among treatments in light transmission were due to the dissolved substances rather than the algal particles themselves (see Supplementary Information). DOC in the generalist (G) treatment was dominated by low-molecular-mass compounds originating from *in situ* primary production²⁸, whereas DOC in the benthic (B) treatment was dominated by larger molecules such as fulvic

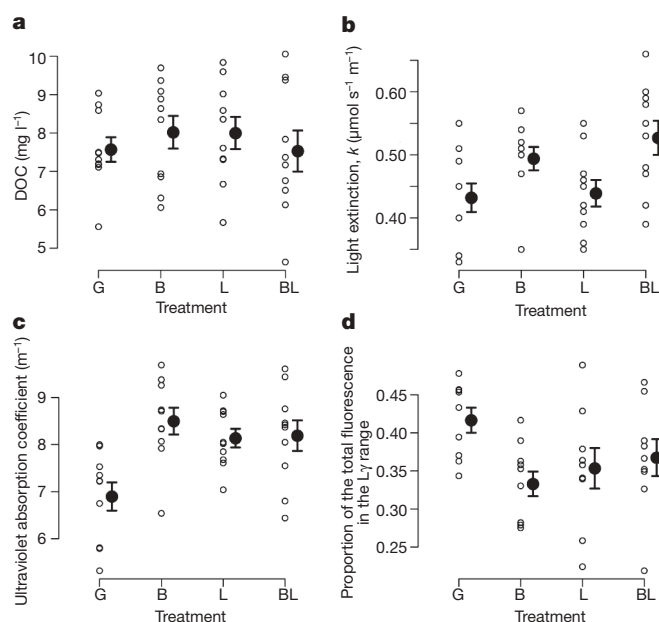


Figure 2 | Dissolved organic content and light transmission across treatments. Small open circles are individual data points; large filled circles are means \pm s.e. ($n = 10$). **a**, Concentration of dissolved organic matter (DOC). **b**, Extinction coefficients of photosynthetically available radiation (PAR, 400–700 nm). **c**, Absorption coefficients for ultraviolet light (A_{320} m^{-1}). In **b** and **c**, lower values indicate higher transparency in the given range of wavelengths. **d**, Ratio of low-molecular-mass molecules (Lγ < 5,000 Da, low excitation wavelengths) to higher molecular mass molecules (Mγ and Hγ, higher excitation wavelengths; see Supplementary Information). A high ratio indicates that the dissolved substances in the water have a smaller size distribution.

acids, which are produced by leaf litter decomposition²⁹ (see Supplementary Information). Because the composition of algal communities is quite sensitive to small changes in the concentration of dissolved substances³⁰, we believe that the differences in light penetration caused by different stickleback predators is relevant for natural lake environments (see Supplementary Information).

This experiment reveals that both specialization and speciation of predators can affect how ecosystems function. Feedbacks between evolutionary diversification and ecosystems are rarely explored but have implications for understanding adaptive radiations. We know that ecology and environmental conditions affect the selective environment for specialization, divergence and speciation⁸. The effect of organisms on their environment through ecosystem engineering could provide a complementary explanation for the tendency of some groups to radiate explosively^{15,23}. Our results illustrate that the effects of stickleback diversification on ecosystem function result from both trophic and non-trophic interactions, and suggest that even short periods of adaptive radiation (~10,000 years) have the potential to transform the structure and functioning of ecosystems. It remains to be seen how the evolution of trophic interactions and ecosystem-engineering effects of predators might shape the evolution of other species in the food web.

METHODS SUMMARY

Mesocosms were constructed from Rubbermaid cattle watering tanks, each 1,136 l, and filled with a mixture of sand and gravel, benthic substrate from natural ponds, well water, and nutrients (Na^+ and PO_4^- ; see Supplementary Information for details). We collected fish from lakes on Texada Island, British Columbia, Canada. Most fish were collected, transported and added to tanks at the beginning of the experiment, but some supplemental fish were collected and transported later in the experiment (see Supplementary Information for details). There was some mortality during the experiment (total deaths = 154). Limnetic fish were more likely to die than benthic fish, although fish biomass did not differ among treatments over the course of the experiment. When dead fish were located, they were removed from the tanks and replaced (see Supplementary Information).

We measured prey community composition by sampling species abundance from the pelagic and benthic zones at the end of the experiment (see Supplementary Information for details). We also took several measures of ecosystem function at regular intervals during the course of the experiment, including chlorophyll *a*, total dissolved phosphorus, nitrates, ammonia, net primary productivity (NPP), gross primary productivity (GPP) and decomposition rates. Finally, we used several techniques to measure and describe the DOC in the tanks, including ultraviolet scanning spectroscopy (UVSS) to measure absorption²⁵, synchronous fluorescence spectroscopy (SFS) to investigate molecular structure²⁶, and fluorescence excitation-emission matrices (FEEM) to assess the relative concentration and potential origins of organic matter²⁹. Full descriptions of all measurements, calculations and data analyses are available in Supplementary Information.

Received 28 October 2008; accepted 12 March 2009.

Published online 1 April 2009.

1. Thompson, J. N. Rapid evolution as an ecological process. *Trends Ecol. Evol.* **13**, 329–332 (1998).
2. Fussmann, G. F., Loreau, M. & Abrams, P. A. Eco-evolutionary dynamics of communities and ecosystems. *Funct. Ecol.* **21**, 465–477 (2007).
3. Yoshida, T., Jones, L. E., Ellner, S. P., Fussmann, G. F. & Hairston, N. G. Jr. Rapid evolution drives ecological dynamics in a predator–prey system. *Nature* **424**, 303–306 (2003).
4. Schmitz, O. J. Predators have large effects on ecosystem properties by changing plant diversity not plant biomass. *Ecology* **87**, 1432–1437 (2006).

5. Loreau, M. *et al.* Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804–808 (2001).
6. McPhail, J. D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): origin of the species pairs. *Can. J. Zool.* **71**, 515–523 (1993).
7. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
8. Seehausen, O. *et al.* Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–626 (2008).
9. Losos, J. B. *Portrait of an Adaptive Radiation: Ecology and Evolution of Anolis Lizards* (Univ. California Press, 2009).
10. Grant, P. R. & Grant, B. R. *How and Why Species Multiply: The Radiation of Darwin's Finches* (Princeton Univ. Press, 2008).
11. Abrams, P. A. The evolution of predator–prey interactions: theory and evidence. *Annu. Rev. Ecol. Syst.* **31**, 79–105 (2000).
12. Palkovacs, E. P. & Post, D. M. Experimental evidence that phenotypic divergence in predators drives community divergence in prey. *Ecology* **90**, 300–305 (2009).
13. Snyder, W. E., Snyder, G. B., Finke, D. L. & Straub, C. S. Predator biodiversity strengthens herbivore suppression. *Ecol. Lett.* **9**, 789–796 (2006).
14. Lennon, J. T. & Martiny, J. B. H. Rapid evolution buffers ecosystem impacts of viruses in a microbial food web. *Ecol. Lett.* **11**, 1178–1188 (2008).
15. Erwin, D. H. Macroevolution of ecosystem engineering, niche construction and diversity. *Trends Ecol. Evol.* **23**, 304–310 (2008).
16. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
17. Rundle, H. D., Nagel, L., Boughman, J. W. & Schluter, D. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306–308 (2000).
18. Boughman, J. W. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411**, 944–947 (2001).
19. Bell, T., Neill, W. E. & Schluter, D. The effect of temporal scale on the outcome of trophic cascade experiments. *Oecologia* **134**, 578–586 (2003).
20. Schmitz, O. J. Effects of predator hunting mode on grassland ecosystem function. *Science* **319**, 952–954 (2008).
21. Goudard, A. & Loreau, M. Non-trophic interactions, biodiversity and ecosystem functioning: an interaction web model. *Am. Nat.* **171**, 91–106 (2008).
22. Stibor, H. *et al.* Copepods act as a switch between alternative trophic cascades in marine pelagic food webs. *Ecol. Lett.* **7**, 321–328 (2004).
23. Jones, C. G., Lawton, J. H. & Shachak, M. Positive and negative effects of organisms as physical ecosystem engineers. *Ecology* **78**, 1946–1957 (1997).
24. Mazumder, A., Taylor, W. D., McQueen, D. J. & Lean, D. R. S. Effects of fish and plankton on lake temperature and mixing depth. *Science* **247**, 312–315 (1990).
25. Williamson, C. E., Morris, D. P., Pace, M. L. & Olson, A. G. Dissolved organic carbon and nutrients as regulators of lake ecosystems: Resurrection of a more integrated paradigm. *Limnol. Oceanogr.* **44**, 795–803 (1999).
26. Retamal, L., Vincent, W. F., Martineau, C. & Osburn, C. L. Comparison of the optical properties of dissolved organic matter in two river-influenced coastal regions of the Canadian arctic. *Estuar. Coast. Shelf Sci.* **72**, 261–272 (2007).
27. Ghan, D., McPhail, J. D. & Hyatt, K. D. The temporal-spatial pattern of vertical migration by the freshwater copepod *Skistodiaptomus oregonensis* relative to predation risk. *Can. J. Fish. Aquat. Sci.* **55**, 1350–1363 (1998).
28. Obernosterer, I. & Benner, R. Competition between biological and photochemical processes in the mineralization of dissolved organic carbon. *Limnol. Oceanogr.* **49**, 117–124 (2004).
29. McKnight, D. M. *et al.* Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* **46**, 38–48 (2001).
30. Pienitz, R. & Vincent, W. F. Effect of climate change relative to ozone depletion on UV exposure in subarctic lakes. *Nature* **404**, 484–487 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Albert, T. Vines, D. Yim, P. Tamkee, J. Courchesne, R. Barrett, K. Marchinko, M. Arnegard, J. Sashaw, J. Gosling, S. Hausch, J. Rosenfeld and S. Rogers for assistance in the laboratory and the field. We thank E. B. Rosenblum and members of the Harmon laboratory for comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.J.H. (lukeh@uidaho.edu).

A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*

Evan Z. Macosko¹, Navin Pokala¹, Evan H. Feinberg¹, Sreekanth H. Chalasani¹, Rebecca A. Butcher², Jon Clardy² & Cornelia I. Bargmann¹

Innate social behaviours emerge from neuronal circuits that interpret sensory information on the basis of an individual's own genotype, sex and experience. The regulated aggregation behaviour of the nematode *Caenorhabditis elegans*, a simple animal with only 302 neurons, is an attractive system to analyse these circuits. Wild social strains of *C. elegans* aggregate in the presence of specific sensory cues, but solitary strains do not^{1–4}. Here we identify the RMG inter/motor neuron as the hub of a regulated circuit that controls aggregation and related behaviours. RMG is the central site of action of the neuropeptide receptor gene *npr-1*, which distinguishes solitary strains (high *npr-1* activity) from wild social strains (low *npr-1* activity); high RMG activity is essential for all aspects of social behaviour. Anatomical gap junctions connect RMG to several classes of sensory neurons known to promote aggregation, and to ASK sensory neurons, which are implicated in male attraction to hermaphrodite pheromones⁵. We find that ASK neurons respond directly to pheromones, and that high RMG activity enhances ASK responses in social strains, causing hermaphrodite attraction to pheromones at concentrations that repel solitary hermaphrodites. The coordination of social behaviours by RMG suggests an anatomical hub-and-spoke model for sensory integration in aggregation, and points to functions for related circuit motifs in the *C. elegans* wiring diagram.

Many naturally isolated social strains of *C. elegans* aggregate into feeding groups with dozens of animals, although other strains, including the laboratory strain N2, are solitary^{1,2}. Aggregating strains show several behavioural changes compared to solitary feeders: they accumulate on the border of a lawn of bacterial food (bordering) and move rapidly on food. Aggregation, bordering and rapid movement are coordinately controlled by the neuropeptide Y receptor homologue NPR-1 (ref. 2). Solitary strains have a high-activity form of NPR-1 (215-valine), whereas aggregating strains have a low-activity form of NPR-1 (215-phenylalanine); *npr-1* null mutants also aggregate^{2,6}. Neuropeptide control of aggregation provides an analogy with mammalian social behaviour, which is regulated by the neuropeptides oxytocin and vasopressin⁷. In addition to genetic regulation by *npr-1*, aggregation is sensitive to environmental signals. It is stimulated by URX sensory neurons that detect environmental oxygen⁴, and ASH and ADL sensory neurons that sense noxious stimuli³. Attraction to low-oxygen environments promotes accumulation at the lawn border and feeding in groups, which have low oxygen levels compared to the open lawn^{4,8}. Population density, food availability³ and environmental stressors⁹ also modulate aggregation. The site of integration of these diverse cues is unknown.

How NPR-1 acts to regulate behaviour is not well understood. A previous report using a genomic *npr-1* fragment identified the oxygen-sensing URX neuron as a site of *npr-1* action, but behavioural rescue was incomplete, with rescue of aggregation, partial rescue of bordering,

and no rescue of rapid movement¹⁰. To identify other neurons in which NPR-1 promotes solitary behaviour, we first established that a full-length *npr-1* complementary DNA expressed from the endogenous *npr-1* promoter rescued solitary behaviour in the strong loss-of-function mutant *npr-1(ad609lf)*, then refined the essential site of expression using other characterized promoters (Fig. 1a, b). Because promoter expression patterns in *C. elegans* can vary between transgenes, we used a bicistronic messenger RNA to express both *npr-1* and green fluorescent protein (GFP), and identified GFP-positive neurons in each rescued line with solitary behaviour (Supplementary Table 1 and Methods). Only promoters driving expression in the inter/motor neuron RMG showed robust rescue of aggregation, bordering, and locomotion speed (Fig. 1b and Supplementary Fig. 1).

We next addressed whether RMG expression of *npr-1* is sufficient to suppress aggregation. No RMG-specific promoter is known, so an intersectional strategy was developed to drive *npr-1* expression only in cells that express both *flp-21* and *ncs-1*, using Cre-mediated recombination between *loxP* sites that flanked transcriptional stop sequences. When *ncs-1::nCre* and *flp-21::LoxStopLox::GFP* strains were crossed together, the intersection between *ncs-1* and *flp-21* allowed strong and consistent GFP expression only in RMG and M2 pharyngeal neurons (Fig. 2a). We next inserted the *npr-1* cDNA into the *flp-21::LoxStopLox* plasmid (Fig. 2a); in *npr-1(lf)* animals expressing both *flp-21::LoxStopLox::npr-1* and *ncs-1::nCre*, aggregation, bordering and high speed on food were strongly suppressed (Fig. 2b). M2 is synaptically isolated from neurons implicated in these behaviours, so we conclude that RMG expression of *npr-1* can block aggregation and related behaviours.

Mammalian neuropeptide Y receptors generally inhibit neurotransmitter release^{11,12}. To determine whether NPR-1 suppresses aggregation by inhibiting or by activating RMG, we killed RMG in wild-type and *npr-1(lf)* animals using a laser microbeam, anticipating an effect on the genotype(s) in which RMG is normally active. Killing RMG in *npr-1(lf)* eliminated aggregation, bordering and rapid movement (Fig. 2c, d), whereas killing RMG in solitary wild-type animals had no effect (Fig. 2d). These results show that RMG neurons stimulate aggregation-related behaviours in *npr-1* mutants, and indicate that NPR-1 inhibits RMG activity in solitary strains.

Inspection of the *C. elegans* wiring diagram¹³ revealed that RMG is the hub of a gap-junction network connecting seven classes of neurons, including the oxygen-sensitive URX neurons and the nociceptive ASH and ADL neurons previously implicated in aggregation behaviour^{3,10} (Fig. 3a). RMG-ablated *npr-1* animals were normal in their avoidance of high osmolarity, a behaviour mediated by ASH¹⁴ (Supplementary Fig. 2). Therefore RMG is not essential for all functions of associated sensory neurons, but selectively required for aggregation and related behaviours.

¹Howard Hughes Medical Institute, Laboratory of Neural Circuits and Behavior, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.

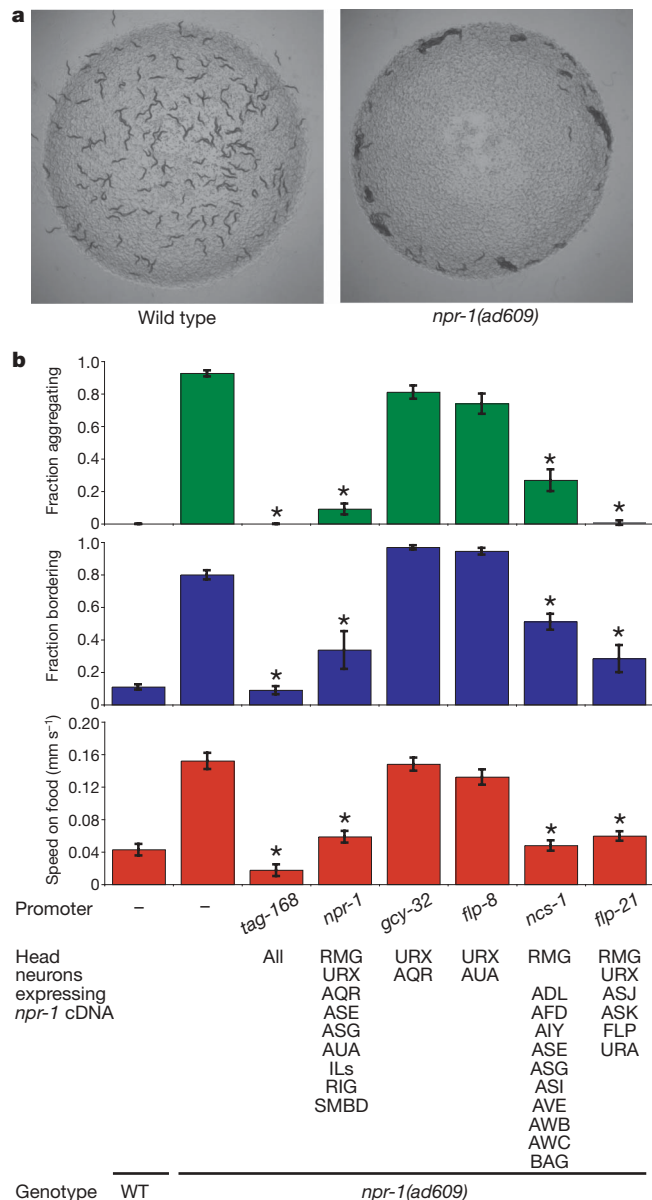


Figure 1 | Selective expression of NPR-1 suppresses aggregation and related behaviours in *npr-1* mutants. **a**, Solitary behaviour of 150 wild type N2 animals (left) and aggregation behaviour of 150 *npr-1(ad609)* animals (right). Magnification $\times 2$. **b**, Behavioural phenotypes of *npr-1(ad609)* animals expressing an *npr-1* cDNA under a pan-neuronal promoter (*tag-168*), its endogenous promoter (*npr-1*), URX promoters (*gcy-32* and *flp-8*) and RMG promoters (*ncs-1* and *flp-21*). For all figures, full promoter expression patterns are in Supplementary Table 1. Error bars indicate s.d. Asterisk, different from *npr-1(ad609)* ($P < 0.01$, Bonferroni test).

Among the other neurons anatomically coupled to RMG, the ASK neurons were of particular interest. ASK is one of several neurons that integrate pheromone and food signals to regulate *C. elegans* development¹⁵, and it has recently been implicated in male attraction to hermaphrodite pheromones⁵. The role of ASK was probed using a *tax-4* mutation that affects sensory transduction: *tax-4* encodes a cyclic GMP-gated transduction channel expressed in ASK and other sensory neurons, but not in RMG¹⁶. *tax-4;npr-1(lf)* double mutants are strongly suppressed for aggregation and related behaviours¹⁰, and rescue of these behaviours requires *tax-4* expression in URX and an unknown sensory neuron¹⁰. We addressed whether ASK might be the unknown neuron. Indeed, the simultaneous expression of *tax-4* in URX and ASK resulted in near-complete rescue of aggregation and related behaviours in *tax-4;npr-1(lf)* double mutants (Fig. 3b).

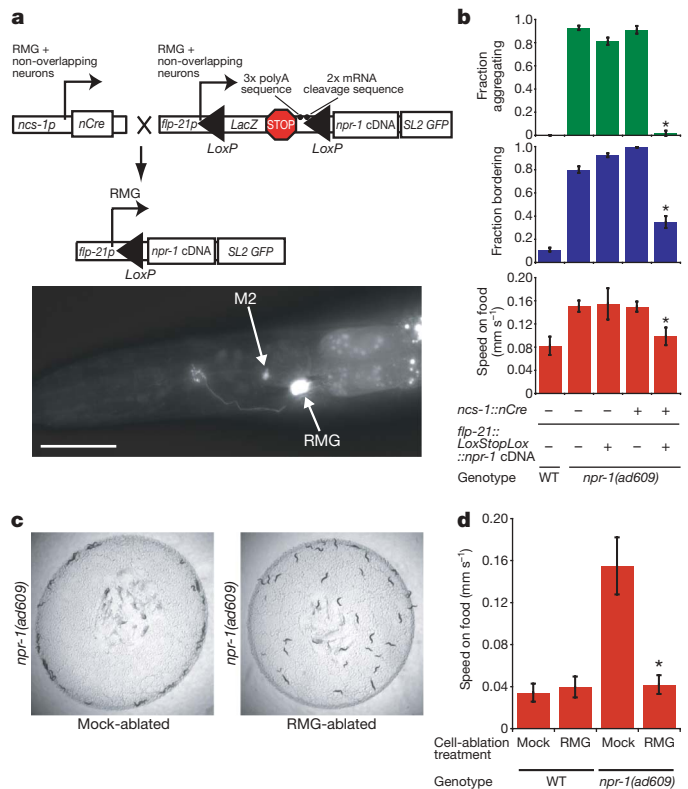


Figure 2 | Inhibition of RMG by NPR-1 suppresses social behaviour. **a**, Top, intersectional Cre/Lox strategy to express *npr-1* specifically in RMG. Bottom, L4 larva expressing *ncs-1::nCre* and *flp-21::LoxStopLox::GFP*. Scale bar, 30 μ m. **b**, Aggregation and related behaviours of *npr-1(ad609)* animals carrying *ncs-1::nCre* and/or *flp-21::LoxStopLox::npr-1* transgenes. Asterisk, different from *npr-1(ad609)* ($P < 0.01$, Student's *t*-test). WT, wild type. **c**, Mock-ablated or RMG-ablated *npr-1(ad609)* animals (mock-ablated: 97.1% bordering, 40% aggregating; RMG-ablated: 17% bordering, 0% aggregating; $\chi^2 = 43.05$, $P < 0.001$). Magnification $\times 1.6$. **d**, Locomotion speed of wild-type and *npr-1(ad609)* animals, mock-ablated or RMG-ablated. Asterisk, different from mock-ablated *npr-1(ad609)* ($P < 0.01$, Student's *t*-test). Error bars indicate s.d.

Rescue was also observed after expression of *tax-4* in URX and ASJ neurons, which synapse onto ASK (Fig. 3b). Thus ASK and ASJ promote aggregation-related behaviours.

The connectivity of RMG suggests two models of behavioural output: RMG could integrate sensory input through gap junctions and stimulate aggregation using its own chemical synapses; or RMG could modify the output of associated sensory neurons, which all have chemical synapses. RMG is presynaptic to head muscles and interneurons that control forward and backward locomotion¹³ (Fig. 3a). To determine whether the synaptic output of RMG promotes aggregation, we used the Cre/Lox system to express the light chain of tetanus toxin (TeTx) in RMG of *npr-1(lf)* mutants. TeTx inhibits synaptic transmission by cleaving the synaptic vesicle protein synaptobrevin¹⁷. Aggregation and related behaviours were partially suppressed by TeTx expression in RMG (Fig. 3c). Coexpression of TeTx with a cleavage-resistant *C. elegans* synaptobrevin (Q68V) mutant¹⁷ significantly suppressed the RMG::TeTx effect, confirming that TeTx acts via synaptobrevin cleavage (Supplementary Fig. 3). Aggregation was also suppressed by the expression of TeTx in ASK and ASJ, further implicating these neurons in aggregation behaviours, and TeTx expression in both RMG and ASK+ASJ neurons had additive effects (Fig. 3c). Silencing or killing URX neurons suppresses aggregation^{10,18}, but TeTx expression in URX neurons had little effect unless the other neurons were silenced (Fig. 3c). These results indicate that synaptic outputs for aggregation are distributed, with contributions from both RMG and ASK+ASJ neurons.

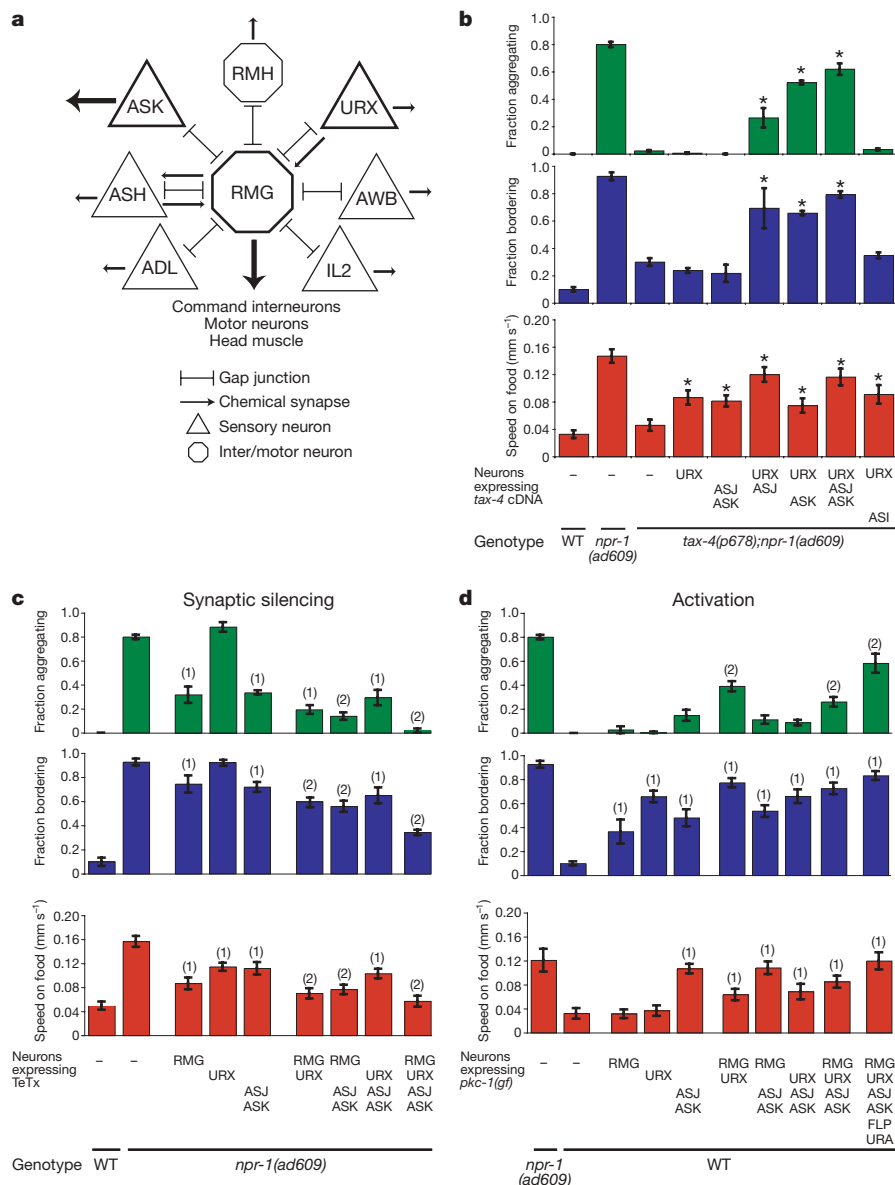


Figure 3 | ASK and ASJ sensory neurons promote aggregation. **a**, Circuit diagram of neurons with gap junctions to RMG. RMG may also form gap junctions with RMF¹³. **b**, Rescue of aggregation and related behaviours in *tax-4(p678);npr-1(ad609)* animals expressing a *tax-4* cDNA. Asterisk, different from *tax-4;npr-1*. **c**, Aggregation and related behaviours of *npr-1(ad609)* animals expressing TeTx. Statistics: (1), different from

We next determined whether activation of RMG or coupled neurons in solitary wild-type animals might induce aggregation. Neurons were activated by expressing a constitutively active protein kinase C homologue of *C. elegans* (*pkc-1(gf)*) that promotes synaptic transmission¹⁹ and neuropeptide release²⁰, and may have further excitatory properties²¹. Expression of *pkc-1(gf)* in most RMG-coupled neurons elicited aggregation, bordering and high speed in solitary strains, a near-complete transformation of their behaviour (Fig. 3d). Expression in subsets of neurons had partial effects, suggesting contributions from RMG, URX, ASK+ASJ and possibly other cells (Fig. 3d). Like the behaviour of *npr-1(lf)* strains, *pkc-1(gf)*-induced behaviours were suppressed by killing RMG (Supplementary Fig. 3b). Thus simultaneous activation of RMG and sensory neurons by *pkc-1(gf)* can drive aggregation and related behaviours.

The dual involvement of ASK in aggregation and male attraction to hermaphrodites⁵ prompted an examination of pheromone responses in aggregating strains. A class of *C. elegans* pheromones termed ascarosides

is attractive to males, but repulsive to solitary hermaphrodites, suggesting a role in sex-specific attraction for mating (Supplementary Fig. 4a)²². Ascarosides are constitutively secreted by *C. elegans*, providing a plausible aggregation signal^{23,24}. Solitary wild-type hermaphrodites were repelled by ascarosides; in contrast, *npr-1(lf)* hermaphrodites were attracted to low levels of ascarosides, with responses resembling those of males (Fig. 4a, b). Expression of *npr-1* in RMG restored pheromone avoidance to *npr-1(lf)* hermaphrodites, linking this behaviour to the RMG circuit (Fig. 4b).

Attraction to pheromones was absent in *tax-4;npr-1(lf)* double mutants, suggesting that *tax-4*-expressing sensory neurons detect ascarosides (Fig. 4c). Rescue of *tax-4* in ASK neurons restored ascaroside attraction to *tax-4;npr-1(lf)* strains, providing evidence that ASK is a relevant pheromone sensor (Fig. 4c). In a systematic analysis of the three pheromone components, a combination of C3 ascaroside with C6 or C9 drove effective attraction via ASK (Supplementary Fig. 4b, c). Developmental effects of ascarosides are greatest in the

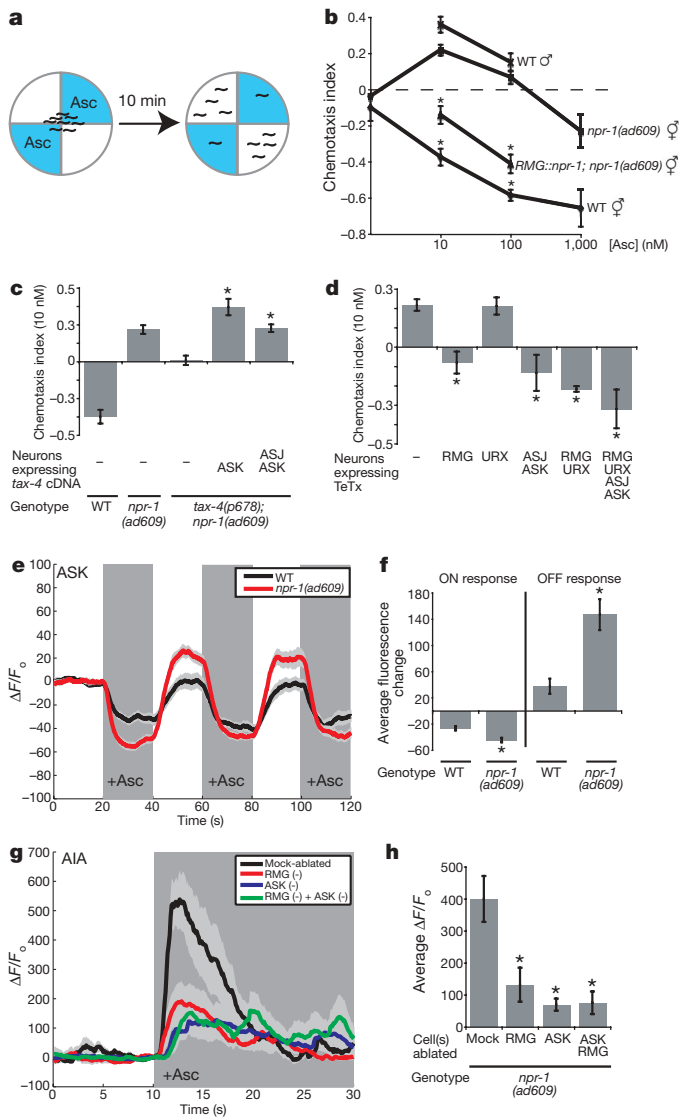


Figure 4 | Behavioural and neuronal responses to pheromones. **a**, Diagram of pheromone chemotaxis assay. Asc, three equimolar ascarosides (C3, C6 and C9). **b**, Ascaroside chemotaxis. Asterisk, different from *npr-1(ad609)*. **c**, ASK expression of *tax-4* restores pheromone attraction to *tax-4;npr-1*. Asterisk, different from *tax-4;npr-1*. **d**, TeTx expression in RMG or ASJ and ASK eliminates pheromone attraction in *npr-1(ad609)*. Asterisk, different from *npr-1(ad609)*. In **b–d**, $P < 0.01$, Bonferroni test. **e**, Ascaroside (100 nM) decreases G-CaMP calcium signals in ASK ($n = 17$ animals each). **f**, Average ASK fluorescence change to first ascaroside addition (ON) and removal (OFF). Asterisks, different from wild type ($P < 0.01$, t -test). **g**, Ascaroside (1 μ M) induces G-CaMP calcium signals in AIA interneurons of *npr-1(ad609)*; for ablations, $n \geq 10$ animals; mock-ablated, $n = 16$. In **e** and **g**, dark shading indicates presence of ascarosides, light shading indicates s.e.m. **h**, Average AIA fluorescence change in the 5 s after ascaroside addition. Asterisk, different from mock-ablated ($P < 0.01$, Bonferroni test). In **b–d**, **f** and **h**, error bars indicate s.e.m.

same combinations, and require higher pheromone concentrations than attraction, suggesting that attraction occurs at physiological pheromone levels²⁴. Inhibiting synaptic transmission from ASK+ASJ or RMG neurons eliminated ascaroside attraction (Fig. 4d). The correlation between cells required for pheromone attraction and aggregation supports the hypothesis that these behaviours are functionally related.

Sensory properties of ASK were examined directly by monitoring sensory-evoked calcium transients with the genetically encoded calcium indicator G-CaMP²⁵. In both wild-type and *npr-1(lf)* animals, ASK responded to ascaroside cocktails (100 pM–1 μ M) with a rapid

diminution of fluorescence suggesting decreased calcium levels; fluorescence recovered after ascaroside removal (Fig. 4e and Supplementary Fig. 4d). The rapid response in ASK neurons supports their identification as ascaroside-sensing neurons; the calcium decrease suggests that ASK uses a hyperpolarizing mode of sensory transduction^{26,27}. At attractive nanomolar ascaroside concentrations, ASK calcium responses were reliably greater in *npr-1* animals than in wild type, with a greater calcium decrease after ascaroside addition and a greater rebound after ascaroside removal (Fig. 4e, f).

To determine whether this apparent change in ASK activity could propagate across synapses, ascaroside responses were monitored in a synaptic target of ASK, the AIA interneuron. Ascaroside cocktails elicited increased G-CaMP fluorescence in AIA, suggesting depolarization (Fig. 4g); the average magnitude of this signal was significantly greater in *npr-1(lf)* than in wild-type animals (Supplementary Fig. 4e). The AIA response was diminished in *npr-1* animals whose ASK neurons were killed with a laser, suggesting that ASK sensory input is an important source of ascaroside signals to AIA (Fig. 4g, h). The inversion of calcium signals (decrease in ASK, increase in AIA) suggests that ASK makes inhibitory synapses onto AIA. Ascaroside-induced AIA calcium signals were also diminished when the RMG neurons were killed (Fig. 4h). An RMG–ASK double ablation resembled an ASK ablation alone, indicating that RMG and ASK affect AIA through a common process (Fig. 4g, h). The imaging results indicate that ASK senses ascarosides (along with other neurons), that the ASK response is propagated to downstream neurons, that RMG enhances ASK signalling, and that high *npr-1* activity diminishes it.

These results provide insight into behavioural mechanisms of aggregation, the anatomical circuit underlying the behaviour, and the regulatory role of *npr-1*. Solitary animals ignore oxygen in the presence of food, and are repelled by ascarosides produced by other animals. In social *npr-1(lf)* animals, oxygen-sensing URX neurons promote accumulation at the lawn border, and ascaroside-sensing ASK neurons promote attraction to other animals (or neutralize repulsion). The altered pheromone response in *npr-1* hermaphrodites demonstrates that aggregation involves directed responses to other nematodes, not just a shared preference for low-oxygen environments. The analysis of RMG suggests a hub-and-spoke model for aggregation behaviour, in which distributed sensory inputs are coordinated through gap junctions with the RMG hub to produce distinct, distributed synaptic outputs. Mechanistically, calcium imaging suggests that RMG amplifies weak sensory signals in ASK to stimulate pheromone attraction. For example, the RMG circuit could depolarize ASK to increase tonic transmitter release at rest, and thereby increase the signal ASK sends when hyperpolarized by ascarosides.

In solitary strains, the neuropeptide receptor NPR-1 inhibits RMG function. In one model, NPR-1 might act by closing RMG gap junctions, gating access of sensory neurons to a shared circuit, but sparing their individual synaptic outputs. This instructive model for gap junction regulation is analogous to the dopaminergic regulation of gap junctions in the mammalian retina, in which gap junctions link rod and cone visual pathways to increase light sensitivity at night; during the day, dopamine inhibits gap junctions to increase spectral and spatial resolution²⁸. Alternatively, NPR-1 could alter RMG excitability, and gap junctions could passively propagate this information from RMG to the sensory neurons to change their properties. In both models, *npr-1* shifts the properties of an entire anatomical circuit by modulating a single neuron. As RMG gap junctions are at present defined purely by anatomical criteria, further experiments are needed to determine whether RMG propagates electrical signals, calcium, cAMP, or other information.

Within the *C. elegans* wiring diagram, gap junction distributions are highly skewed. Most neurons have only a few gap junctions, but 17 classes of neurons are gap junction hubs that link seven or more neuronal classes²⁹. We suggest that this circuit motif performs a characteristic computation wherever it appears.

METHODS SUMMARY

Aggregation and bordering behaviours were measured essentially as described²; values report the average fraction of three or more behavioural assays of 150 animals each. Average locomotion speed was calculated by tracking 20 animals for 10 min with an automated tracking system³⁰.

For RMG-selective expression of transgenes, a *LoxP*-flanked *LacZ* sequence containing a transcriptional stop, three repeated poly(A) sequences, and two repeated mRNA cleavage sequences, was inserted upstream of *npr-1::SL2::GFP* under the control of the *flp-21* promoter (*flp-21::LoxStopLox::cDNA[GFP, npr-1, TeTx, or pkc-1(gf)]*). Transgenic animals containing this plasmid were crossed with animals expressing *nCre* under the *ncs-1* promoter (*ncs-1::nCre*). Strong and consistent expression was observed in RMG and M2 neurons; ADL, ASJ and ASK were seen weakly and inconsistently.

For ascaroside chemotaxis assays, washed animals were placed in the centre of a four-quadrant plate with ascarosides in alternating quadrants, and scored after 10 min. A chemotaxis index was calculated as (the number of animals on pheromone quadrants – the number of animals on buffer quadrants)/(total number of animals). In the cartoon in Fig. 4a, the chemotaxis index = –0.6. In Fig. 4 a cocktail of three ascarosides was used; individual ascarosides and other combinations are shown in Supplementary Fig. 4.

Calcium imaging of the AIA and ASK neurons was performed in a custom-fabricated microfluidic device, essentially as described²⁶. For ASK imaging, the transgene *kyEx2866* was used, with *GCaMP2.2b* (gift from L. Looger) expressed under the *sra-9* promoter. For AIA imaging, the transgene *kyEx2916* was used, with *GCaMP2.2b* expressed under the T01A4.2 promoter. ASK fluorescence was recorded in the neuronal cell body, and AIA fluorescence was measured in the dorsal AIA process in the nerve ring.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 October 2008; accepted 11 February 2009.

Published online 6 April 2009.

- Hodgkin, J. & Doniach, T. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics* **146**, 149–164 (1997).
- de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94**, 679–689 (1998).
- de Bono, M., Tobin, D. M., Davis, M. W., Avery, L. & Bargmann, C. I. Social feeding in *Caenorhabditis elegans* is induced by neurons that detect aversive stimuli. *Nature* **419**, 899–903 (2002).
- Gray, J. M. *et al.* Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430**, 317–322 (2004).
- Srinivasan, J. *et al.* A blend of small molecules regulates both mating and development in *Caenorhabditis elegans*. *Nature* **454**, 1115–1118 (2008).
- Rogers, C. *et al.* Inhibition of *Caenorhabditis elegans* social feeding by FMRFamide-related peptide activation of NPR-1. *Nature Neurosci.* **6**, 1178–1185 (2003).
- Hammock, E. A. & Young, L. J. Oxytocin, vasopressin and pair bonding: implications for autism. *Phil. Trans. R. Soc. Lond. B* **361**, 2187–2198 (2006).
- Cheung, B. H., Cohen, M., Rogers, C., Albayram, O. & de Bono, M. Experience-dependent modulation of *C. elegans* behavior by ambient oxygen. *Curr. Biol.* **15**, 905–917 (2005).
- Davies, A. G., Bettinger, J. C., Thiele, T. R., Judy, M. E. & McIntire, S. L. Natural variation in the *npr-1* gene modifies ethanol responses of wild strains of *C. elegans*. *Neuron* **42**, 731–743 (2004).
- Coates, J. C. & de Bono, M. Antagonistic pathways in neurons exposed to body fluid regulate social feeding in *Caenorhabditis elegans*. *Nature* **419**, 925–929 (2002).
- Plummer, M. R., Rittenhouse, A., Kanevsky, M. & Hess, P. Neurotransmitter modulation of calcium channels in rat sympathetic neurons. *J. Neurosci.* **11**, 2339–2348 (1991).
- Toth, P. T., Bindokas, V. P., Bleakman, D., Colmers, W. F. & Miller, R. J. Mechanism of presynaptic inhibition of neuropeptide Y at sympathetic nerve terminals. *Nature* **364**, 635–639 (1993).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Bargmann, C. I., Thomas, J. H. & Horvitz, H. R. Chemosensory cell function in the behavior and development of *Caenorhabditis elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **55**, 529–538 (1990).
- Schackwitz, W. S., Inoue, T. & Thomas, J. H. Chemosensory neurons function in parallel to mediate a pheromone response in *C. elegans*. *Neuron* **17**, 719–728 (1996).
- Komatsu, H., Mori, I., Rhee, J. S., Akaike, N. & Ohshima, Y. Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. *Neuron* **17**, 707–718 (1996).
- Schiavo, G. *et al.* Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359**, 832–835 (1992).
- Chang, A. J., Chronis, N., Karow, D. S., Marletta, M. A. & Bargmann, C. I. A distributed chemosensory circuit for oxygen preference in *C. elegans*. *PLoS Biol.* **4**, e274 (2006).
- Sieburth, D. *et al.* Systematic analysis of genes required for synapse structure and function. *Nature* **436**, 510–517 (2005).
- Sieburth, D., Madison, J. M. & Kaplan, J. M. PKC-1 regulates secretion of neuropeptides. *Nature Neurosci.* **10**, 49–57 (2007).
- Okochi, Y., Kimura, K. D., Ohta, A. & Mori, I. Diverse regulation of sensory signaling by *C. elegans* nPKC-epsilon/eta TTX-4. *EMBO J.* **24**, 2127–2137 (2005).
- White, J. Q. *et al.* The sensory circuitry for sexual attraction in *C. elegans* males. *Curr. Biol.* **17**, 1847–1857 (2007).
- Butcher, R. A., Fujita, M., Schroeder, F. C. & Clardy, J. Small-molecule pheromones that control dauer development in *Caenorhabditis elegans*. *Nature Chem. Biol.* **3**, 420–422 (2007).
- Butcher, R. A., Ragains, J. R., Kim, E. & Clardy, J. A potent dauer pheromone component in *Caenorhabditis elegans* acts synergistically with other components. *Proc. Natl Acad. Sci. USA* **105**, 14288–14292 (2008).
- Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nature Biotechnol.* **19**, 137–141 (2001).
- Chalasani, S. H. *et al.* Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nature* **450**, 63–70 (2007).
- Suzuki, H. *et al.* Functional asymmetry in *Caenorhabditis elegans* taste neurons and its computational role in chemotaxis. *Nature* **454**, 114–117 (2008).
- Ribelayga, C., Cao, Y. & Mangel, S. C. The circadian clock in the retina controls rod-cone coupling. *Neuron* **59**, 790–801 (2008).
- Chen, B. L., Hall, D. H. & Chklovskii, D. B. Wiring optimization can relate neuronal structure and function. *Proc. Natl Acad. Sci. USA* **103**, 4723–4728 (2006).
- Ramot, D., Johnson, B. E., Berry, T. L. Jr, Carnell, L. & Goodman, M. B. The parallel worm tracker: a platform for measuring average speed and drug-induced paralysis in nematodes. *PLoS One* **3**, e2208 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Looger for *GCaMP2.2b*, M. Nonet for cleavage-resistant synaptobrevin, and J. Ragains for synthesizing ascarosides. This work was funded by the Howard Hughes Medical Institute, the Harold and Leila Y Mathers Charitable Foundation, the Jensam Foundation, and National Institute of Health grants GM07739 (E.Z.M. and E.H.F.), CA24487 (J.C.) and GM077943 (R.A.B.). C.I.B. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions E.Z.M. performed experiments; N.P., E.H.F., S.C., R.A.B. and J.C. developed experimental methods and reagents; E.Z.M. and C.I.B. designed and interpreted experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.I.B. (cori@rockefeller.edu).

METHODS

Strains were grown and maintained under standard conditions³¹. Standard methods were used for molecular biology. A complete strain list and information on clones and transgenes is included in Supplementary Information.

Behavioural analysis. Aggregation and bordering behaviours were on NGM plates seeded with 200 μ l OP50 for 48 h and 150 adult animals per assay². Behaviours were scored 2 h after picking the animals to the assay plates. An animal was scored as aggregating if it touched two or more other animals, and bordering if it was less than 2 mm from the edge of the bacterial lawn. To determine the rate of locomotion on food, 10-cm NGM agar plates were seeded with a thin layer of OP50 bacteria and allowed to grow overnight at 25 °C. Filter paper soaked in 20 mM copper chloride was placed around the perimeter of the plate to prevent animals from escaping from view. Twenty animals were then picked to the centre of the plate, allowed to rest for 1 h, and recorded for 10 min. Average speed was calculated from speed values determined every ten seconds by an automated worm tracker (<http://wormsense.stanford.edu/tracker>)³⁰.

A previous report indicated that social behaviour could be partially suppressed by expressing *npr-1* in URX neurons under *gcy-32* or *flp-8* promoters¹⁰. We did not observe strong suppression using any of the three tested URX-expressing promoters: *gcy-32*, *flp-8* or *gcy-35*. One possible explanation for this incongruity is that the published rescue uses an intron-containing genomic fragment of *npr-1*, whereas we used an *npr-1* cDNA. Some introns in *C. elegans* act as tissue-specific or nonspecific transcriptional enhancers³². Therefore the *npr-1* genomic constructs may have directed some expression in RMG; furthermore, some fragments of the *flp-8* promoter, which was more potent in the published report, are occasionally expressed in RMG³³. An alternative explanation is that the published genomic clone may result in higher *npr-1* expression in URX than the cDNA clone. Social behaviour can be partially suppressed by silencing or killing URX^{10,34} and the reported partial suppression is consistent with silenced URX neurons. Whether the higher or lower levels of *npr-1* are more physiological is unclear, but it is a concern that transgenes can be toxic when expressed at high levels, even when they drive 'inert' molecules such as GFP. In any case, the effects of RMG *npr-1* expression described here are much stronger than the reported effects of URX *npr-1* expression.

Chemotaxis to pheromone was assayed in 10-cm Petri plates divided into four quadrants³⁵ with ascarosides mixed into the agar in alternating quadrants. For each assay, 200 worms were picked to a seeded plate for 2–3 h. The worms were then washed three times with chemotaxis buffer³⁶, and placed in the centre of the assay plate.

Osmotic avoidance was measured using the dry drop test³⁷. Eight to ten animals per condition were placed on individual NGM plates without food. After 10 min, a small drop of 1 M glycerol, dispensed from a micropipette, was placed in the path of a forward-moving animal, so that the drop soaked into the agar as the animal reached the drop. A reversal away from the glycerol was scored as an avoidance event. Each animal was tested 8–10 times, from which a fractional response was calculated. The avoidance index is the average fractional response for all animals tested in a given condition.

Laser ablation. For behavioural assays, the RMG inter/motor neurons were targeted for laser killing using a transgenic strain expressing GFP in RMG (*ncs-1::GFP*) and a MicroPoint laser system. Cells were ablated at the L1 stage

as described³⁸, and behaviour was assayed approximately 72 h later. Individual ablated animals were examined for the absence of RMG fluorescence to confirm cell death. For calcium imaging experiments, neurons were identified using Nomarski optics on the basis of their position and morphology, then killed with a laser. Adult animals were visually scored for aggregation-related behaviours to confirm the death of RMG.

Calcium imaging. Calcium imaging of the AIA and ASK neurons was performed in a custom-fabricated polydimethylsiloxane chamber that physically immobilized animals with their nose in a stream of fluid²⁶. Ascaroside stimuli were presented by switching the fluid streams that bathed the nose. G-CaMP fluorescence was recorded by a Nikon CoolSnap camera, regions of interest were tracked, and fluorescence intensity analysed using Matlab scripts.

The near-ultraviolet-light wavelengths used to excite the G-CaMP fluorophore elicited fluorescence changes in ASK; therefore, as described for other light-sensitive neurons³⁹, the ultraviolet-response was desensitized by a 2–3 min light pre-exposure before the beginning of all experiments monitoring ASK and AIA activity. Previous studies are consistent with a possible light-sensitive function of ASK⁴⁰, and with a demonstrated light-sensitive activity of the synaptically connected ASJ neurons⁴⁰. All imaging experiments used a 1:1:1 ratio of three ascarosides (Supplementary Fig. 4a), each at the same concentration (100 pM–1 μ M). Behavioural responses to pheromones were examined in strains used for imaging, and resembled those of the parental *npr-1* strains.

To quantify ASK responses, the average response in the first 20 s after ascaroside addition in each recording was subtracted from the average of 10 s before ascaroside addition (ON response), and the average response in the first 20 s after ascaroside removal was subtracted from the average of 10 s before ascaroside removal (OFF response) (Fig. 4f). To quantify AIA responses, the average change in fluorescence during the first 5 s after presentation of the ascaroside stimulus was calculated for each recording (Fig. 4h).

31. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
32. Hobert, O. et al. Regulation of interneuron function in the *C. elegans* thermoregulatory pathway by the *ttx-3* LIM homeobox gene. *Neuron* **19**, 345–357 (1997).
33. Kim, K. & Li, C. Expression and regulation of an FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans*. *J. Comp. Neurol.* **475**, 540–550 (2004).
34. Sweeney, S. T., Broadie, K., Keane, J., Niemann, H. & O'Kane, C. J. Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects. *Neuron* **14**, 341–351 (1995).
35. Hart, A. C. et al. *The Wormbook: Behavior* <http://www.wormbook.org/chapters/www_behavior/behavior.html> (2006).
36. Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* **7**, 729–742 (1991).
37. Hilliard, M. A., Bargmann, C. I. & Bazzicalupo, P. C. *C. elegans* responds to chemical repellents by integrating sensory inputs from the head and the tail. *Curr. Biol.* **12**, 730–734 (2002).
38. Bargmann, C. I. & Avery, L. Laser killing of cells in *Caenorhabditis elegans*. *Methods Cell Biol.* **48**, 225–250 (1995).
39. Hilliard, M. A. et al. *In vivo* imaging of *C. elegans* ASH neurons: cellular response and adaptation to chemical repellents. *EMBO J.* **24**, 63–72 (2005).
40. Ward, A., Liu, J., Feng, Z. & Xu, X. Z. Light-sensitive neurons and channels mediate phototaxis in *C. elegans*. *Nature Neurosci.* **11**, 916–922 (2008).

LETTERS

Toxin B is essential for virulence of *Clostridium difficile*

Dena Lyras¹, Jennifer R. O'Connor^{1,3}, Pauline M. Howarth¹, Susan P. Sambol³, Glen P. Carter¹, Tongted Phumoonna¹, Rachael Poon^{1,2}, Vicki Adams¹, Gayatri Vedantam³, Stuart Johnson³, Dale N. Gerding³ & Julian I. Rood^{1,2}

Clostridium difficile is the leading cause of infectious diarrhoea in hospitals worldwide, because of its virulence, spore-forming ability and persistence^{1,2}. *C. difficile*-associated diseases are induced by antibiotic treatment or disruption of the normal gastrointestinal flora^{3,4}. Recently, morbidity and mortality resulting from *C. difficile*-associated diseases have increased significantly due to changes in the virulence of the causative strains and antibiotic usage patterns^{1,2,5,6}. Since 2002, epidemic toxinotype III NAP1/027 strains^{1,2}, which produce high levels of the major virulence factors, toxin A and toxin B, have emerged. These toxins have 63% amino acid sequence similarity⁷ and are members of the large clostridial glucosylating toxin family, which are monoglucosyltransferases that are pro-inflammatory, cytotoxic and enterotoxic in the human colon^{8–10}. Inside host cells, both toxins catalyse the transfer of glucose onto the Rho family of GTPases, leading to cell death^{8,11}. However, the role of these toxins in the context of a *C. difficile* infection is unknown. Here we describe the construction of isogenic *tcdA* and *tcdB* (encoding toxin A and B, respectively) mutants of a virulent *C. difficile* strain and their use in the hamster disease model to show that toxin B is a key virulence determinant. Previous studies showed that purified toxin A alone can induce most of the pathology observed after infection of hamsters with *C. difficile*^{8,9,12} and that toxin B is not toxic in animals unless it is co-administered with toxin A, suggesting that the toxins act synergistically¹². Our work provides evidence that toxin B, not toxin A, is essential for virulence. Furthermore, it is clear that the importance of these toxins in the context of infection cannot be predicted exclusively from studies using purified toxins, reinforcing the importance of using the natural infection process to dissect the role of toxins in disease.

C. difficile is difficult to manipulate genetically; reproducible methods for the construction of chromosomal mutants have only recently become available^{13,14}. We have developed a method that involves a plasmid, designated the recombination vector, which is unstable in *C. difficile*; this plasmid carries DNA homologous to part of the genome and becomes inserted at that genomic location, resulting in a specific gene disruption. Using this process, we have successfully mutated two regulatory genes¹³ and others have generated mutations in additional genes¹⁵. This methodology has now made possible studies aimed at dissecting the role of toxin A and toxin B in *C. difficile*-associated diseases (CDAD) by constructing and analysing isogenic strains of this bacterium.

Toxin A and toxin B are encoded by the *tcdA* and *tcdB* genes, respectively, located within a 19.6-kilobase (kb) pathogenicity locus, the PaLoc⁹ (Fig. 1a), which encodes three other genes involved in toxin production^{16,17}. To construct the plasmids required for the targeted inactivation of *tcdA* and *tcdB*, 566-base-pair (bp) *tcdA*- or

tcdB-specific polymerase chain reaction (PCR) products were cloned into pJIR1456. These PCR products corresponded to an essential region of the amino-terminal domain between the DXD catalytic motif and amino acid 546 (Fig. 1b)¹⁸. Insertion of these plasmids into the corresponding chromosomal gene region by homologous recombination should lead to disruption of the gene, resulting in a truncation of the toxins (at leucine residues 492 and 493 in toxins A and B, respectively), thereby attenuating toxin function.

The *tcdA*- and *tcdB*-specific recombination vectors were transferred to JIR8094 (ref. 13), a derivative of *C. difficile* strain 630, by RP4-mediated conjugative transfer from *Escherichia coli*. Two putative independently derived toxin A (*tcdA*Δ1 and *tcdA*Δ2) and five putative independently derived toxin B (including *tcdB*Δ1 and *tcdB*Δ2) mutants were isolated. Note that because each of these

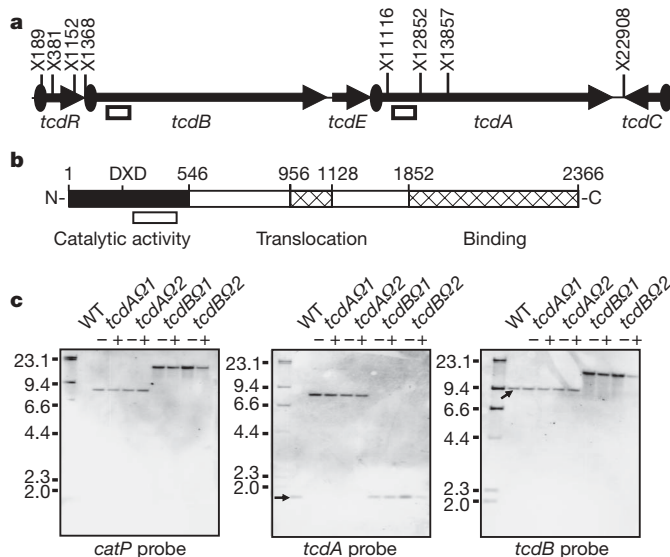


Figure 1 | Genetic organization of the *C. difficile* PaLoc and analysis of toxin mutants. **a**, PaLoc genetic organization. *Xba*I sites (X) are shown, with the PaLoc coordinate indicated. Boxes indicate regions used in the recombination vectors and the *tcdB*- and *tcdA*-specific probes. Ovals show the gene-specific promoters and arrows the direction of transcription. **b**, Structural organization of toxin B (toxin A is similar). Coordinates indicate the amino acid number. Relevant domains are shown¹⁸. The box indicates the region in the recombination vectors. **c**, Southern hybridization of mutants. Genomic DNA was *Xba*I-digested and blots probed as shown. λ -HindIII markers are indicated (kb). WT represents strain JIR8094; the mutant lanes are indicated. DNA was purified from cultures grown in the absence (–) or presence (+) of thiamphenicol selection.

¹Australian Bacterial Pathogenesis Program and ²Australian Research Council Centre of Excellence in Structural and Functional Microbial Genomics, Department of Microbiology, Monash University, Victoria 3800, Australia. ³Medical Service and Research Service, Hines VA Hospital, Hines, Illinois 60141, USA, and Infectious Disease Section, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois 60153, USA.

mutants was derived from a separate conjugation experiment, the mutants are truly independent. The genotypes of the mutants were confirmed by PCR analysis (data not shown) and Southern blotting (Fig. 1c). The insertion of the recombination vectors was shown to be specific and to have disrupted the respective *tcdA* and *tcdB* genes, with the wild-type bands replaced by bands corresponding to insertion of the recombination vectors. For each specific mutant the other toxin gene had a profile identical to wild type. Even though the insertions were generated by single crossovers, and were therefore reversible¹³, the resultant chromosomal mutants were very stable; no evidence of deletion of the inserted plasmids was detected in the Southern blots (Fig. 1c). Autonomously replicating plasmids could not be extracted from these strains, suggesting that once chromosomal insertion had occurred, the plasmid was maintained in an inserted state, as previously found¹³.

Western blots using toxin-A-specific antibodies showed that the *tcdA* mutants no longer produced toxin A and had a similar profile to an avirulent strain that does not carry the PaLoc (Fig. 2a). By comparison, the *tcdB* mutants produced toxin A, although they reproducibly appeared to make more toxin A than the wild type. This finding was confirmed by quantitative real-time reverse transcriptase PCR (qRT-PCR), which showed that the *tcdB* mutants expressed *tcdA* at 2.2–3.3-fold higher levels than the wild type. The basis for this increased expression of toxin A is not known. qRT-PCR showed that no differences in expression of the other PaLoc genes (*tcdR*, *tcdE* and *tcdC*) were observed in any of the mutants (data not shown).

Cytotoxicity assays using HT29 cells, which are more sensitive to toxin A than other cell lines^{19,20}, showed that supernatants from the toxin A mutants were significantly less toxic ($P = 0.021$) than supernatants from the wild type (Fig. 2b). These assays also showed that toxin A produced by the wild-type strain and by the *tcdB* mutants was cytotoxic and therefore functional (Fig. 2b). The latter activity was neutralized by toxin-A-specific antibodies (Fig. 2c), verifying that the observed toxicity resulted from the action of toxin A. Vero cell cytotoxicity assays were also carried out on these supernatants; the cytotoxic activity observed is predominantly toxin-B-mediated^{8,19}. Consequently, the *tcdA* mutants showed the same cytotoxic phenotype as the wild-type strain (Fig. 2d). These cytopathic effects were neutralized by toxin-B-specific antibodies,

confirming that toxin B was responsible for the observed Vero cell toxicity (Fig. 2e). By contrast, supernatants from the *tcdB* mutants were significantly less cytotoxic ($P = 0.0058$) for Vero cells than wild-type supernatants. These mutants showed a similar phenotype to the avirulent strain although they did have low levels of toxicity (Fig. 2d), consistent with the relatively lower cytopathic effects of toxin A on Vero cells. This low level of cytotoxicity was not neutralized by pre-incubation with toxin B antibodies (Fig. 2e).

All of the genetic and phenotypic assays used to analyse the mutant strains therefore showed that each pair of mutants, although isolated independently, behaved identically. Attempts were made to complement these mutations with the wild-type genes, but these experiments were not successful. The intact genes are very difficult to clone into shuttle vectors because they are very large (8.1 kb and 7.1 kb for *tcdA* and *tcdB*, respectively) and contain repeat regions^{21,22}. For this reason, in subsequent virulence experiments we used two independently isolated mutants for each gene, to rule out the possibility that any observed phenotypic effects were the result of other mutations.

The hamster disease model parallels most of the recognized features of human CDAD, especially with respect to the susceptibility of the animal to *C. difficile* infection after the administration of antimicrobial agents. Because it is currently the best animal model available²³ it was used to compare the virulence of the *C. difficile* mutants.

On challenge with the wild type, all ten infected hamsters were colonized approximately 2 days after inoculation and 9 out of 10 animals died approximately 5 days after colonization (Fig. 3). The two independent toxin A mutants, which produced wild-type levels of toxin B *in vitro*, yielded similar results, with 9 out of 10 and 8 out of 10 hamsters being colonized after challenge with the *tcdA*Q1 and *tcdA*Q2 mutants, respectively. The mean number of days (\pm s.d.) to colonization was similar after inoculation with wild-type (2.1 ± 1.2), *tcdA*Q1 (1.9 ± 0.3) and *tcdA*Q2 (1.9 ± 1.1) strains. These results matched *in vitro* studies that investigated the adherence of these strains to polarized Caco-2 human intestinal epithelial cells under anaerobic conditions, where no significant differences in adherence were observed (data not shown).

The mean time from colonization to death was similar for hamsters colonized with the wild-type strain (9 out of 10) and the *tcdA*Q1 (9 out

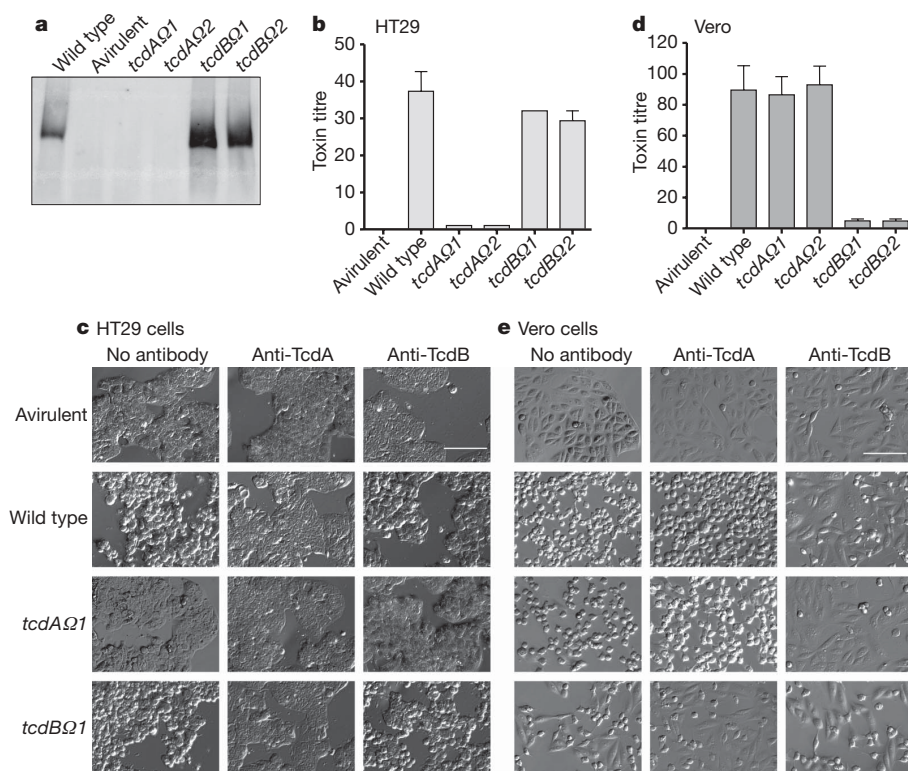


Figure 2 | Comparative analysis of toxin production by wild-type and mutant *C. difficile* strains. The wild type is JIR8094 and the avirulent strain is CD37, which does not produce toxins²⁹. **a**, Western blot using toxin-A-specific antibodies. **b**, **d**, Toxin A and toxin B cytotoxicity assays using HT-29 and Vero cells, respectively. Data represent the mean \pm s.e.m.; $n = 3$. **c**, **e**, Neutralization of toxicity. Culture supernatants were added to the cells and morphological changes observed after 24 h (No antibody) or pre-treated with toxin A (Anti-TcdA) and toxin B (Anti-TcdB) neutralizing antibodies before their addition to the cells. Representative images are shown. No neutralization was observed when supernatants were pre-treated with anti-ErmB methylase antibodies (data not shown). Bar indicates 1 mm.

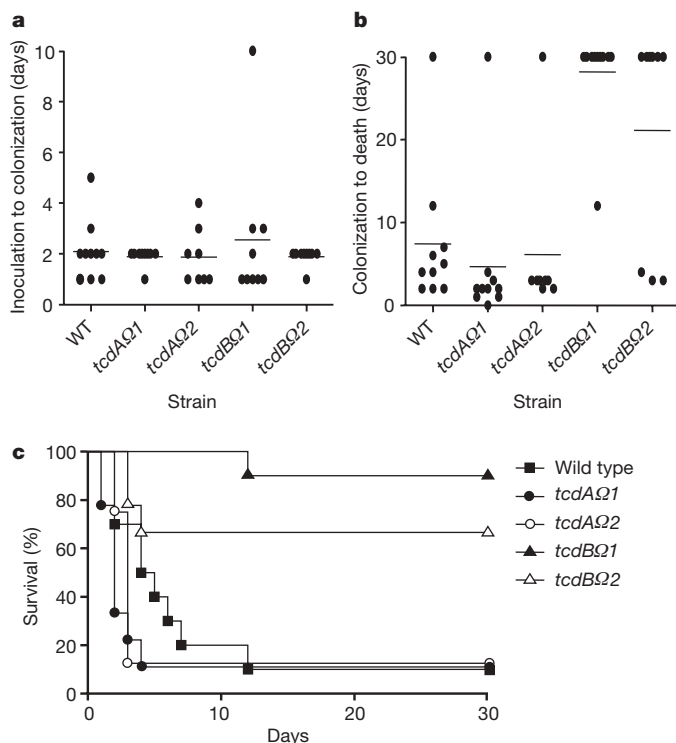


Figure 3 | Virulence of *C. difficile* wild-type and mutant strains in hamsters. **a**, Days from inoculation of *C. difficile* to colonization of clindamycin-treated hamsters. The rate of colonization was 10 out of 10 for wild type; 9 out of 10 for *tcdA*Q1; 8 out of 10 for *tcdA*Q2; 9 out of 10 for *tcdB*Q1; 9 out of 10 for *tcdB*Q2. **b**, Days from colonization to death. All hamsters that survived were killed at day 30. **c**, Kaplan-Meier survival estimate demonstrating days from colonization with *C. difficile* strains to death. $n = 10$.

of 9) and *tcdA*Q2 (7 out of 8) mutants (Fig. 3). Faecal pellets from hamsters infected with the *tcdA*Q2 mutant were analysed for plasmid stability; toxin analysis was also performed on these specimens. All of the colonies ($n = 800$) recovered from the eight *tcdA*Q2 colonized hamsters were thiamphenicol resistant, confirming that the inserted recombination vector was retained in these isolates. Seven of the eight colonized hamsters died (Fig. 3c), the one surviving animal lost colonization after only 2 days. Confirmatory PCRs, with primers specific for the gene disruption, were performed on a minimum of four post-colonization isolates (from different hamsters) for each group of animals; the results showed that they all carried an insertionally inactivated *tcdA* gene.

All of the faecal pellets isolated from *tcdA*Q2-infected hamsters were strongly cytotoxic (using a cytotoxin assay that primarily measures toxin B), indicating the presence of active toxin B, and were negative for toxin A in an enzyme immunoassay, as expected. Overall, 16 out of 17 (94%) of the hamsters colonized with the toxin A mutants died, providing clear evidence that the *tcdA* mutants were still virulent even though they no longer produced toxin A, either *in vitro* or *in vivo*. These data provide clear evidence that toxin A is not essential for virulence in the hamster model, at least for a derivative of strain 630.

By contrast, the virulence trials involving the toxin B mutants demonstrated that toxin B was required for virulence (Fig. 3). For each of the *tcdB* mutants tested, 9 out of 10 hamsters were colonized after challenge. The time from challenge to colonization (*tcdB*Q1, 2.6 ± 2.9 days; *tcdB*Q2, 1.9 ± 0.3 days) was similar to that observed for the wild type and the toxin A mutants. However, only 4 out of 19 of the hamsters (21%) colonized with the toxin B mutants died (Fig. 3c), which was statistically different ($P = 0.001$) from the wild-type strain (10 out of 10 deaths). The hamster that died after challenge with the *tcdB*Q1 mutant died 12 days after colonization, although the interval

from colonization to death in the three hamsters that died after challenge with the *tcdB*Q2 mutant was similar to that of the wild type. Plasmid stability analysis and toxin testing of faecal pellets from the *tcdB*Q2 trial explained these results. All of the colonies ($n = 600$) recovered from the six surviving hamsters were thiamphenicol resistant, indicating that the integrated recombination vector was carried by these strains. Faecal pellets from 5 out of 6 surviving hamsters were enzyme immunoassay (EIA) positive for toxin A and were either cytotoxin negative ($n = 2$) or weakly cytotoxic ($n = 4$), consistent with the absence of toxin B and the relatively lower cytotoxicity of toxin A. These results provided evidence that the isolates recovered from the surviving hamsters were no longer virulent, even though they continued to produce toxin A *in vivo*. Of the three hamsters that died, colonies from two animals ($n = 200$) were all thiamphenicol resistant. However, only 74% of the isolates from the third hamster were thiamphenicol resistant. Restriction endonuclease analysis (REA) typing of two of the thiamphenicol-sensitive isolates confirmed that they were derived from the wild-type strain, suggesting that they represented revertants that had lost the integrated plasmid and therefore displayed a wild-type (toxin A⁺B⁺) phenotype. In addition, faecal pellets from all three hamsters that died were strongly cytotoxic (toxin B positive) and toxin A positive by EIA. We conclude that the thiamphenicol-sensitive isolates were revertants and were able to produce toxin B, reinforcing the requirement for toxin B in the virulence of *C. difficile*. We suggest that the two hamsters that carried thiamphenicol-resistant strains, but whose faecal isolates were toxin A⁺B⁺, had mixed infections of the mutant strain and a revertant strain capable of producing wild-type toxin B at a level that was not detected by our analysis.

Hamsters challenged with the toxin A mutants were as likely to die as those infected with the wild-type strain (Fig. 2a), and animals challenged with the toxin B mutants were much more likely to survive (Fig. 3c), even though these strains produce more toxin A than the wild type. Among the few hamsters that died after challenge with the toxin B mutants, our analysis suggested that a proportion of the *C. difficile* population infecting those animals produced toxin B. These results support the conclusion that the production of toxin B, not toxin A, correlates with a virulent *C. difficile* phenotype.

In conclusion, four independently derived toxin A or toxin B mutants of *C. difficile* were constructed and characterized *in vitro* and *in vivo* using the hamster CDAD model. The results demonstrated that toxin B is an essential virulence factor, as disruption of the *tcdB* gene led to a significantly attenuated virulence phenotype. Isolates that produced toxin B but not toxin A retained a wild-type virulence phenotype, which provides further confirmation of the paramount role of toxin B in disease pathogenesis. Unexpectedly, the presence of toxin A in the absence of toxin B, even at levels higher than in the wild-type strain, was not lethal in the hamster model. Although this finding ostensibly contradicts a study which showed that antibodies to toxin A protect toxoid A-immunized hamsters from disease²⁴, vaccination studies do not provide direct evidence of the role of a bacterial component in disease. Therefore, it is concluded that toxin A is not an essential virulence determinant in a derivative of the virulent *C. difficile* strain 630, although it may still have a non-essential role in the disease process.

Some previous studies have suggested that toxin A may not be the major virulence factor of *C. difficile* and that toxin B might not be dependent on toxin A for function^{23,25}; however, a lack of genetic tools did not allow these hypotheses to be tested. Other studies suggested that toxin B was unable to initiate disease unless toxin A was present^{9,12}. The data generated from our study do not support the hypothesis that both toxin A and toxin B are required to cause disease. Clinical variants expressing toxin B but not toxin A are being isolated with increasing frequency²⁶ and disease resulting from these strains is similar to that following infection with strains producing both toxins²⁷. The observation that clinical isolates that lack toxin A are still capable of causing severe disease reinforces the proposition that toxin A is not essential for virulence and that toxin B does not

require the presence of toxin A for its *in vivo* activity^{9,23,26,27}. Note that naturally occurring toxin A⁺B⁻ isolates of *C. difficile* have not been reported. Earlier studies on toxin A and toxin B function and their putative cooperativity involved purified toxins or culture supernatants^{9,12}. The disparity between these studies and those described here accentuates the importance of using the natural infection process to dissect the role of toxins in disease pathogenesis and reinforces the critical need for the use of isogenic strains in such studies.

The clinical diagnosis of *C. difficile* predominantly involved immunoassays for the detection of toxin A until the discovery of clinical disease caused by toxin A⁻B⁺ isolates that were not detected by these assays. It is now evident from our data that in the absence of culturing for the organism, the diagnostic emphasis should focus on the detection of both toxins by EIA and/or a standard cytotoxin assay that measures primarily toxin B²⁶, especially with toxin-A-deficient strains being isolated with increasing frequency. Rapid and accurate diagnosis is of critical importance because of the changing epidemiology and increased virulence of *C. difficile* strains^{1,2}, and because this organism is now causing disease in individuals previously thought not to be at risk, such as children and pregnant women²⁸. It is clear that our findings represent a paradigm shift in our understanding of the pathogenesis of *C. difficile*-mediated disease and will have major implications for the diagnosis, control and treatment of these syndromes.

METHODS SUMMARY

The toxin-A- and toxin-B-specific recombination vectors pJIR3051 and pJIR3050 were constructed by cloning a 566-bp *tcdA* (GenBank accession M30307) PCR product (nucleotides 1071–1637) and the equivalent *tcdB* (GenBank accession X53138) region, respectively, into the shuttle vector pJIR1456 (GenBank accession U90554). These plasmids were transferred by conjugation to *C. difficile* strain JIR8094 and the resultant chromosomal mutants analysed by PCR, Southern blotting and plasmid rescue¹³.

Filtered supernatants of 2TY broth cultures (72 h) were used for western blots, cytotoxicity and toxin neutralization assays. Serial twofold dilutions of these supernatants were used in Vero and HT-29 cell cytotoxicity assays, with the end-point scored as the last dilution yielding complete cytopathic effects (CPE). Toxins were neutralized with affinity-purified polyclonal antibodies specific for toxin A and toxin B; neutralization was carried out at the last dilution giving complete CPE. Partially purified ammonium-sulphate-precipitated toxin proteins were western blotted using toxin-A-specific affinity-purified polyclonal goat antibodies.

Syrian hamsters ($n = 10$) were infected with *C. difficile* spores for virulence testing²³. Clindamycin was administered orally to render the hamsters susceptible to infection (day zero). On day five hamsters received 1,000 colony-forming units (c.f.u.) of spores by gastric inoculation. Faecal pellets were collected daily for 12 days, then weekly until day 30 and inoculated onto selective TCCFA plates to monitor colonization. Colonies were typed by REA to confirm the identity of the infecting isolate. Faecal pellets were analysed for the presence of toxin A using a toxin-A-specific enzyme-linked immunosorbent assay (ELISA) and for the presence of toxin B using a human foreskin fibroblast cytotoxicity assay.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 June 2008; accepted 29 January 2009.

Published online 1 March 2009.

- McDonald, L. C. *et al.* An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N. Engl. J. Med.* **353**, 2433–2441 (2005).
- Warny, M. *et al.* Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* **366**, 1079–1084 (2005).
- Bartlett, J. G. *et al.* Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. *Gastroenterology* **75**, 778–782 (1978).
- Bartlett, J. G. Antibiotic-associated diarrhea. *N. Engl. J. Med.* **346**, 334–339 (2002).
- Redelings, M. D., Sorvillo, F. & Mascola, L. Increase in *Clostridium difficile*-related mortality rates, United States, 1999–2004. *Emerg. Infect. Dis.* **13**, 1417–1419 (2007).
- O'Brien, J. A., Lahue, B. J., Caro, J. J. & Davidson, D. M. The emerging infectious challenge of *Clostridium difficile*-associated disease in Massachusetts hospitals: clinical and economic consequences. *Infect. Control Hosp. Epidemiol.* **28**, 1219–1227 (2007).
- von Eichel-Streiber, C. *et al.* Comparative sequence analysis of the *Clostridium difficile* toxins A and B. *Mol. Gen. Genet.* **233**, 260–268 (1992).

- Just, I. & Gerhard, R. Large clostridial cytotoxins. *Rev. Physiol. Biochem. Pharmacol.* **152**, 23–47 (2004).
- Voth, D. E. & Ballard, J. D. *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin. Microbiol. Rev.* **18**, 247–263 (2005).
- Jank, T. & Aktories, K. Structure and mode of action of clostridial glucosylating toxins: the ABCD model. *Trends Microbiol.* **16**, 222–229 (2008).
- Just, I. *et al.* Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* **375**, 500–503 (1995).
- Lyerly, D. M., Saum, K. E., MacDonald, D. K. & Wilkins, T. D. Effects of *Clostridium difficile* toxins given intragastrically to animals. *Infect. Immun.* **47**, 349–352 (1985).
- O'Connor, J. R. *et al.* Construction and analysis of chromosomal *Clostridium difficile* mutants. *Mol. Microbiol.* **61**, 1335–1351 (2006).
- Heap, J. T. *et al.* The Clostron: a universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* **70**, 452–464 (2007).
- Dineen, S., Villapakkam, A. C., Nordman, J. T. & Sonenshein, A. L. Repression of *Clostridium difficile* toxin gene expression by CodY. *Mol. Microbiol.* **66**, 206–219 (2007).
- Mani, N. *et al.* Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. *J. Bacteriol.* **184**, 5971–5978 (2002).
- Matamouros, S., England, P. & Dupuy, B. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol. Microbiol.* **64**, 1274–1288 (2007).
- Schirmer, J. & Aktories, K. Large clostridial cytotoxins: cellular biology of Rho/Ras-glucosylating toxins. *Biochim. Biophys. Acta* **1673**, 66–74 (2004).
- Torres, J., Camorlinga-Ponce, M. & Munoz, O. Sensitivity in culture of epithelial cells from rhesus monkey kidney and human colon carcinoma to toxins A and B from *Clostridium difficile*. *Toxicon* **30**, 419–426 (1992).
- Giesemann, T. *et al.* Cholesterol-dependent pore formation of *Clostridium difficile* toxin A. *J. Biol. Chem.* **281**, 10808–10815 (2006).
- Burger, S. *et al.* Expression of recombinant *Clostridium difficile* toxin A using the *Bacillus megaterium* system. *Biochem. Biophys. Res. Commun.* **307**, 584–588 (2003).
- Tang-Feldman, Y. J. *et al.* One-step cloning and expression of *Clostridium difficile* toxin B gene (*tcdB*). *Mol. Cell. Probes* **16**, 179–183 (2002).
- Sambol, S. P. *et al.* Infection of hamsters with epidemiologically important strains of *Clostridium difficile*. *J. Infect. Dis.* **183**, 1760–1766 (2001).
- Kim, P. H., Iaconis, J. P. & Rolfe, R. D. Immunization of adult hamsters against *Clostridium difficile*-associated ileocectitis and transfer of protection to infant hamsters. *Infect. Immun.* **55**, 2984–2992 (1987).
- Du, T. & Alfa, M. J. Translocation of *Clostridium difficile* toxin B across polarized Caco-2 cell monolayers is enhanced by toxin A. *Can. J. Infect. Dis.* **15**, 83–88 (2004).
- Drudy, D., Fanning, S. & Kyne, L. Toxin A-negative, toxin B-positive *Clostridium difficile*. *Int. J. Infect. Dis.* **11**, 5–10 (2007).
- Johnson, S. *et al.* Fatal pseudomembranous colitis associated with a variant *Clostridium difficile* strain not detected by toxin A immunoassay. *Ann. Intern. Med.* **135**, 434–438 (2001).
- Benson, L., Song, X., Campos, J. & Singh, N. Changing epidemiology of *Clostridium difficile*-associated disease in children. *Infect. Control Hosp. Epidemiol.* **28**, 1233–1235 (2007).
- Smith, C. J., Markowitz, S. M. & Macrina, F. L. Transferable tetracycline resistance in *Clostridium difficile*. *Antimicrob. Agents Chemother.* **19**, 997–1003 (1981).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Research at Monash University was supported by Program Grant 284214 from the Australian National Health and Medical Research Council, funding from the ARC Centre of Excellence in Structural and Functional Microbial Genomics and grant A1057637 from the United States National Institute of Allergy and Infectious Diseases. S.J., D.N.G. and G.V. were supported by Merit Review Grants from the United States Department of Veterans Affairs Research Service. We thank D. Lyerly, K. Aktories and C. von-Eichel-Streiber for providing toxin-A-specific and toxin-B-specific antibodies, K. Nagaro and A. Cheknis for assistance with the hamster experiments, V. K. Viswanathan for providing intestinal epithelial cell lines, E. Hartland for providing the HT29 cell line, and M. Merrigan for adherence assays.

Author Contributions D.L., J.R.O'C., G.P.C., V.A. and J.I.R. designed the genetic and molecular aspects of the research, which were carried out by D.L., P.M.H. and G.P.C. S.J., S.P.S., J.R.O'C. and D.N.G. planned and developed the animal experiments, which were carried out by S.P.S. and J.R.O'C. Tissue culture assays were designed, planned and carried out by D.L., G.P.C., R.P., T.P. and G.V. All authors were involved in data analysis and interpretation. The manuscript was primarily written by D.L. and J.I.R. but all authors had very significant input into its content and reviewed and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.I.R. (julian.rood@med.monash.edu.au).

METHODS

Bacterial strains, plasmids and growth conditions. All bacteriological culture media were obtained from Oxoid or Difco. For mating procedures and extraction of plasmid and genomic DNA, *C. difficile* strains were cultured in BHIS broth²⁹, supplemented with 0.1% L-cysteine HCl and 0.375% glucose, or BHIS agar supplemented with 0.09% FeSO₄, Cefoxitin and thiamphenicol (Sigma) were used at 25 µg ml⁻¹ and 10 µg ml⁻¹, respectively. *C. difficile* cultures were grown in an atmosphere of 10% H₂, 10% CO₂ and 80% N₂ at 37 °C in a Coy anaerobic chamber. *E. coli* strains were derivatives of either DH5α (Life Technologies) or S17-1 (ref. 30). *E. coli* strains were cultured aerobically at 37 °C in 2×YT agar or broth³¹. Where appropriate, *E. coli* cultures were supplemented with chloramphenicol, erythromycin or ampicillin, at 30 µg ml⁻¹, 150 µg ml⁻¹, 100 µg ml⁻¹, respectively.

Isolation and manipulation of nucleic acids. Plasmid DNA was isolated from *E. coli* strains grown overnight in 5 ml of broth, with appropriate antibiotic selection, using Qiaprep spin miniprep columns (Qiagen), following the manufacturer's instructions. *E. coli* transformations were performed using the Pipes method³². *C. difficile* genomic DNA was prepared as previously described¹³. Standard methods were used for the digestion, modification, ligation and analysis of plasmid and genomic DNA and PCR products³³. Oligonucleotide primers used in this study are listed in Supplementary Information. Nucleotide sequencing reactions were carried out using a PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems), according to the manufacturer's instructions. Sequence detection was carried out on an Applied Biosystems 3730S Genetic Analyser by Micromon at Monash University and sequences were analysed using Sequencher 3.1 software (Gene Codes Corporation).

qRT-PCR analysis of PaLoc gene expression. The procedure was carried out essentially as before³⁴. In brief, RNA was extracted from 40 ml of *C. difficile* cultures grown to an OD₆₀₀ of approximately 0.3 for *tcdC* gene expression analysis, and from 10 ml of *C. difficile* cultures grown to an OD₆₀₀ of approximately 1.8 for all other PaLoc gene expression analysis, using the Ribopure Bacteria kit (Ambion), according to the manufacturer's instructions. To remove contaminating genomic DNA, the purified RNA was treated with the DNA-free kit (Ambion), according to the manufacturer's instructions. Reverse transcription reactions were then performed using 2 µg of purified RNA, 0.5 µM primer, and the Omniscript Reverse Transcription Kit (Qiagen) according to the manufacturer's instructions. Prior to real-time PCR analysis all RT reactions were diluted twofold and fivefold. Reactions were then performed in a final volume of 25 µl with SYBR Green PCR master mix (Applied Biosystems), 2 µl of diluted RT reactions and 120 nM primers using an Eppendorf Realplex⁴ Mastercycler. Triplicate reactions were performed in multiple experiments using RNA from three biological replicates. The data were normalized to *C. difficile rpoA* RNA levels. Primers used for qRT-PCR are listed in Supplementary Information.

Construction of toxin-A- and toxin-B-specific recombination vectors. Oligonucleotide primers JRP2342 and JRP2343 were used to amplify the 566-bp region corresponding to nucleotides 1071–1637 of *tcdA* (accession number M30307). The corresponding region of *tcdB* (accession number X53138) was amplified using JRP2344 and JRP2345. Because EcoRI restriction sites had been incorporated into the primers, this enzyme was used to clone the products into the corresponding site of the shuttle vector pJIR1456³⁵ (accession number U90554) to construct the *tcdA*- and *tcdB*-targeted recombination vectors pJIR3051 and pJIR3050, respectively.

Isolation of chromosomal *C. difficile* mutants. To construct *C. difficile* strains with mutations resulting from insertion of the recombination vectors in the *tcdA* or *tcdB* genes, the vectors pJIR3051 and pJIR3050 were introduced into *E. coli* strain S17-1, respectively. Strain S17-1 carries a chromosomal copy of the broad-host-range plasmid RP4 and although it is not able to self-transfer it provides all the functions required to mobilize co-resident Inc.P *oriT* plasmids. The *oriT*-containing plasmids pJIR3051 and pJIR3050 were therefore transferred by conjugation to *C. difficile* strain JIR8094, as previously described¹³. In the final step, the growth from each non-selective plate was resuspended in 1 ml of BHI diluent (0.37% BHI, 0.01% sodium thioglycollate) and 100 µl spread onto BHIS agar supplemented with thiamphenicol and cefoxitin, the latter antibiotic being used as a counter selection against *E. coli*. The plates were incubated anaerobically at 37 °C for at least 72 h, until colonies were observed. The colonies then were subcultured onto the same medium and incubated for a further 2 to 3 days. Subcultures that showed bacterial growth were subcultured twice more, then analysed by PCR, Southern blotting and plasmid rescue experiments.

Southern hybridization analysis of *C. difficile* strains. Genomic *C. difficile* DNA was digested with XbaI, subjected to agarose gel electrophoresis and then transferred to a Nylon H+ Hybond membrane (Amersham)³³. Southern hybridization analysis was then carried out on the resultant membrane using

standard methods³⁶. DNA probes were digoxigenin (DIG)-labelled by random PCR labelling, following the manufacturer's instructions (Roche). The blots were hybridized with probes specific for *tcdA* (the 566-bp PCR product of primers JRP2342/JRP2343), *tcdB* (the 566-bp JRP2344/JRP2345 PCR product) and *catP* (0.73-kb product amplified from pJIR1456 using primers JRP2142/JRP2143¹³).

Partial purification of toxins from culture supernatants. The toxins were partially purified by ammonium sulphate precipitation from culture supernatants. Briefly, *C. difficile* was grown overnight in 20 ml of 2TY broth (3.0% tryptone, 2.0% yeast extract and 0.1% sodium thioglycollate) and approximately 3 ml transferred to 90 ml of the same medium. The cultures were then grown for 3 days, after which time the cells were pelleted by centrifugation (10,000g, 15 min, room temperature). The supernatants were then filter sterilized (0.45-µm filters) and solid ammonium sulphate was added with stirring at room temperature, to reach 70% saturation, followed by incubation for 20 min at 4 °C. The precipitated protein was collected and dissolved in 2 ml of water and dialysed overnight at 4 °C in PBS, with four changes of buffer.

Toxin-A-specific western blots. Partially purified toxin proteins were subjected to SDS-PAGE and transferred by electrophoresis to nitrocellulose membranes (Whatman), as previously described³⁴. The membranes were incubated for 1 h in PBS containing 5% dry milk powder and then incubated overnight with affinity purified polyclonal goat antibodies specific for *C. difficile* toxin A (Techlab). Toxin-A-bound antibodies were then detected with peroxidase-conjugated anti-goat antibodies (Chemicon) and the Western Lightning Chemiluminescence reagent kit (Perkin-Elmer), according to the manufacturer's instructions.

Vero and HT-29 cell cytotoxicity assays. To prepare the supernatants used in the toxin assays, *C. difficile* was grown as before in 90 ml of TY broth for 3 days, and the cells pelleted by centrifugation as before. The supernatants were filter sterilized and stored on ice before use. Vero and HT-29 cells were cultured in minimum essential medium (MEM alpha medium: GIBCO, Invitrogen) or McCoy's medium (5A medium modified: GIBCO, Invitrogen), respectively, containing 10% heat-inactivated fetal calf serum (FCS), 100 units ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin in culture flasks at 37 °C in 5% CO₂. The cells were grown to a confluent (Vero) or semi-confluent (HT-29) monolayer and subcultured by incubation in 1 to 2 ml of 0.1% trypsin in 1 mM EDTA. The cells were counted and resuspended in fresh medium at a concentration of 0.25 × 10⁵ (Vero) or 5.0 × 10⁵ (HT-29) cells per ml. One millilitre of the cell suspension was seeded into each well of 24-well plates. The plates were incubated for 20–24 h and the culture medium removed, after which cells were washed with PBS. Serial twofold dilutions of the *C. difficile* culture supernatants were made in PBS and 100 µl added to each well, followed by 400 µl of MEM or McCoy's medium containing 1% heat-inactivated FCS. Negative controls were treated with 500 µl of fresh medium. The plates were incubated at 37 °C in 5% CO₂. The morphological changes were observed by microscopy after 24 h. The cytopathic effect (CPE) was determined on a scale from 0 to +4 in comparison to the negative control wells. The end point was scored as the last dilution at which 100% or 4+ CPE was observed. The assays were performed in triplicate on independent culture supernatants. An Olympus 1X71 inverted microscope was used to visualize the cells at 10× and 20× magnifications.

Toxin neutralization assays. For neutralization of toxin activity, affinity purified polyclonal antibodies specific to toxin A (goat anti-TcdA, Techlab) and toxin B (rabbit anti-TcdB, tgcBIOMICS) were used. Neutralization assays were performed by incubating filtered supernatants with appropriately diluted antibodies for 90 min at room temperature before the addition of the treated supernatant to Vero or HT-29 cell monolayers, together with growth medium supplemented with 1% HI FCS. As a negative control, culture supernatants were treated with a polyclonal antibody reactive against the ErmB RNA methylase.

Anaerobic *C. difficile* adherence assays. Quantification of *C. difficile* adhesion to host cells was achieved by an anaerobic bacterial adherence assay that used a derivative of the Caco-2 human intestinal epithelial cell-line, Caco-2BBE (C2BBE). C2BBE host cells express the human sodium-glucose transporter SGLT1 and elaborate a brush-border when cultured in monolayers. C2BBE cells were cultured in high-glucose (25 mM) Dulbecco's Minimal Eagle Medium (DMEM), 10% fetal bovine serum, 20 mM HEPES, 100 units ml⁻¹ penicillin, and 100 µg ml⁻¹ streptomycin at 37 °C in the presence of 5% CO₂. Cells between passages 25 and 45 were grown as confluent monolayers (1.2 × 10⁶ cells) in 6-well plates, and transferred to antibiotic and serum-free DMEM 24 h before adherence assays. For the assays, C2BBE plates were introduced into the anaerobic chamber just before use, serum-free medium was removed and exponentially growing *C. difficile* strains applied at a multiplicity of infection of 20 in a volume of 2 ml. All bacterial strains used were washed and resuspended in completely anaerobic DMEM with 25 mM CaCl₂ before incubation with host cells. Adherence was allowed to proceed under anaerobic conditions for 40 min. Host cells and adherent bacteria were then washed twice with 1 ml of completely

anaerobic phosphate-buffered saline (PBS), scraped, serially diluted and plated to enumerate adherent *C. difficile*. Each experiment was performed in quadruplicate, and repeated at least three times in its entirety. The percent adherence was calculated as the ratio of recovered *C. difficile* to input *C. difficile* multiplied by 100. The SPSS software package was used for statistical analyses. Significance ($P < 0.01$) was determined using ANOVA to enable comparison between multiple groups of continuous numerical data. The Least Significant Difference test was used for post hoc analyses.

Hamster experiments. *C. difficile* strains were tested in groups of 10 specific pathogen-free, 90–100 g Syrian hamsters. To minimize contamination, hamsters were housed in individual isolator cages fitted with air filters on their lids. Cages, food, water bottles, water and bedding were autoclaved, and personnel handling the animals wore disposable gowns, masks, hats, shoe covers and gloves. Each experimental group was housed in a separate room to prevent cross-contamination with different strains of *C. difficile*^{23,37}.

To prepare spores, *C. difficile* strains were streaked heavily and grown anaerobically at 37 °C on anaerobic blood agar plates (BBL anaerobic blood agar plates, Columbia base, Fisher L21928) to confluency, with an incubation time of 5 to 6 days to induce sporulation. The organisms were then harvested with disposable loops, placed into 10 ml of PBS, washed in PBS, and heat-shocked at 56 °C for 10 min to kill surviving vegetative cells. The spores were centrifuged and resuspended in DMEM, aliquoted, and frozen at –80 °C. The frozen spores were then quantified before use by plating tenfold serial dilutions of the spores onto taurocholate-fructose-agar plates. The agar base was taurocholate-cefoxitin-cycloserine-fructose-agar (TCCFA), but without cycloserine and cefoxitin²³. Spores were diluted in DMEM before inoculation.

For each test strain, hamsters were given clindamycin orogastrically (30 mg kg^{–1}), to render them susceptible to *C. difficile* infection (day zero). On study day five hamsters received 1,000 c.f.u. of *C. difficile* spores by gastric inoculation. Hamsters were then monitored twice daily for 5 days and daily thereafter. Animal bedding was changed and faecal pellets were collected daily. Specimens were inoculated onto selective TCCFA plates and incubated anaerobically at 37 °C to determine if they were colonized with *C. difficile*. Faecal pellets were collected daily for 12 days, then weekly until termination of the study (up to 30 days). *C. difficile* isolates from at least three colonized hamsters per group were typed by restriction endonuclease analysis (REA) to confirm the identity of the infecting *C. difficile* isolate³⁸. Hamster groups were assessed for colonization, time interval between challenge with *C. difficile* and colonization and time interval between colonization and death. Note that this study was approved by the Institutional Animal Care and Use Committee (IACUC) at the Hines VA Hospital.

The wild-type parent strain JIR8094 and two independently derived toxin A (*tcdA*Δ1 and *tcdA*Δ2, JIR8253 and JIR8263, respectively) and toxin B mutants (*tcdB*Δ1 and *tcdB*Δ2, JIR8253 and JIR8263, respectively) were each analysed in groups of 10 hamsters. During the hamster experiments with the second group of independent mutants, faecal pellets from all of the *C. difficile*-colonized hamsters

were analysed for stability of the toxin A or toxin B mutation. Faecal samples taken after colonization were homogenized in 3.75 ml of PBS and serially diluted and subcultured onto TCCFA. For each hamster, 100 *C. difficile* colonies were then individually patched onto BHIS agar containing cefoxitin and thiampenicol, to confirm that the *catP* antibiotic resistance marker from the recombination plasmid was still present, and onto BHIS agar containing cefoxitin as a positive control for growth. Hamster faecal pellets, homogenized in 0.75 ml PBS, were also analysed for the presence of toxin A, using a toxin-A-specific ELISA (TechLab) and for the presence of toxin B using a human foreskin fibroblast cytotoxicity assay (human foreskin fibroblast cell monolayers, Diagnostic Hybrids; anti-*C. difficile* toxin, Trinity Biotech).

For the toxin A ELISA, 0.2 g of homogenized faecal pellet sample was added to 0.2 ml of diluent and mixed. The samples were clarified by centrifugation and the assay performed according to the manufacturer's instructions. The OD₄₅₀ was determined using a microtitre plate reader (Bio-Rad). Samples were considered positive if their OD₄₅₀ was at least 0.19, and negative if their OD₄₅₀ was less than 0.05; there were no intermediate results. For the toxin B cytotoxicity assay, 0.1 g of homogenized faecal sample was mixed with 0.5 ml of diluent and mixed. The samples were clarified by centrifugation, filtered through 0.45 µm pore filters (Millipore), and the assay performed according to the manufacturer's directions. The cytotoxic effect was recorded at 24 h and all positive results were confirmed by neutralization with antitoxin. Results were recorded as negative (–), weak cytotoxicity (+; less than 10% cell rounding), or high cytotoxicity (+++; greater than 80% cell rounding); there were no intermediate results. REA typing was performed and analysed as previously described³⁷.

30. Simon, R., Priefer, U. & Puhler, A. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. *Bio/Technology* **1**, 784–790 (1983).
31. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, 1972).
32. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).
33. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
34. Carter, G. P. et al. Binary toxin production in *Clostridium difficile* is regulated by CdtR, a LytTR family response regulator. *J. Bacteriol.* **189**, 7290–7301 (2007).
35. Lyras, D. & Rood, J. I. Conjugative transfer of RP4-oriT shuttle vectors from *Escherichia coli* to *Clostridium perfringens*. *Plasmid* **39**, 160–164 (1998).
36. Lyras, D. & Rood, J. Transposition of Tn4451 and Tn4453 involves a circular intermediate that forms a promoter for the large resolvase, TnpX. *Mol. Microbiol.* **38**, 588–601 (2000).
37. Razaq, N. et al. Infection of hamsters with historical and epidemic BI types of *Clostridium difficile*. *J. Infect. Dis.* **196**, 1813–1819 (2007).
38. Clabots, C. R. et al. Development of a rapid and efficient restriction endonuclease analysis typing system for *Clostridium difficile* and correlation with other typing systems. *J. Clin. Microbiol.* **31**, 1870–1875 (1993).

LETTERS

Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation

Myriam Aouadi^{1*}, Gregory J. Tesz^{1*}, Sarah M. Nicoloso¹, Mengxi Wang¹, My Chouinard¹, Ernesto Soto¹, Gary R. Ostroff¹ & Michael P. Czech¹

Gene silencing by double-stranded RNA, denoted RNA interference, represents a new paradigm for rational drug design¹. However, the transformative therapeutic potential of short interfering RNA (siRNA) has been stymied by a key obstacle—safe delivery to specified target cells *in vivo*². Macrophages are particularly attractive targets for RNA interference therapy because they promote pathogenic inflammatory responses in diseases such as rheumatoid arthritis, atherosclerosis, inflammatory bowel disease and diabetes³. Here we report the engineering of β 1,3-D-glucan-encapsulated siRNA particles (GeRPs) as efficient oral delivery vehicles that potentially silence genes in mouse macrophages *in vitro* and *in vivo*. Oral gavage of mice with GeRPs containing as little as $20 \mu\text{g kg}^{-1}$ siRNA directed against tumour necrosis factor α (Tnf- α) depleted its messenger RNA in macrophages recovered from the peritoneum, spleen, liver and lung, and lowered serum Tnf- α levels. Screening with GeRPs for inflammation genes revealed that the mitogen-activated protein kinase kinase kinase 4 (Map4k4) is a previously unknown mediator of cytokine expression. Importantly, silencing Map4k4 in macrophages *in vivo* protected mice from lipopolysaccharide-induced lethality by inhibiting Tnf- α and interleukin-1 β production. This technology defines a new strategy for oral delivery of siRNA to attenuate inflammatory responses in human disease.

To accomplish oral delivery of siRNA to macrophages in mice, we targeted M cells in intestinal wall Peyer's patches (Supplementary Fig. 1) to transfer micrometre-sized β 1,3-D-glucan particles to the underlying gut-associated lymphatic tissue (GALT)⁴. Phagocytosis by macrophages and dendritic cells in the GALT occurs by means of the dectin-1 (also known as CLEC7A) receptor and perhaps other β 1,3-D-glucan-receptor-mediated pathways⁵. Evidence suggests that GALT macrophages may traffic away from the gut and infiltrate other reticuloendothelial system tissues, such that over time some of the total body macrophages contain ingested β 1,3-D-glucan particles⁶. We prepared hollow, porous micrometre-sized shells composed primarily of β 1,3-D-glucan by treating baker's yeast with a series of alkaline, acid and solvent extractions to remove cytoplasm and other cell wall polysaccharides⁷ (Fig. 1). Such empty β 1,3-D-glucan shells are about 2–4 μm in diameter, and can be fluorescently labelled for tracking. The anionic siRNA within GeRPs is bound between cationic polyethylenimine (PEI) layers through electrostatic interactions in a pH-dependent manner. On phagocytosis of GeRPs by macrophages, the acidic pH in phagosomes promotes siRNA release through the porous GeRP wall.

Preliminary experiments using two siRNA oligonucleotides against Tnf- α within GeRPs showed extensive phagocytosis of GeRPs by the primary macrophages and gene silencing (Supplementary Fig. 2c, d). We then screened in peritoneal exudate cell macrophages (PECs) for

candidate intracellular signalling proteins that might control Tnf- α expression. One of these was the Map4k4, a germinal centre protein kinase that facilitates Tnf- α signalling itself^{8–10}. β 1,3-D-glucan shells derivatized with a green fluorescein probe were loaded with scrambled or Map4k4 siRNA (oligo 1 in Supplementary Table 1) coupled to a red

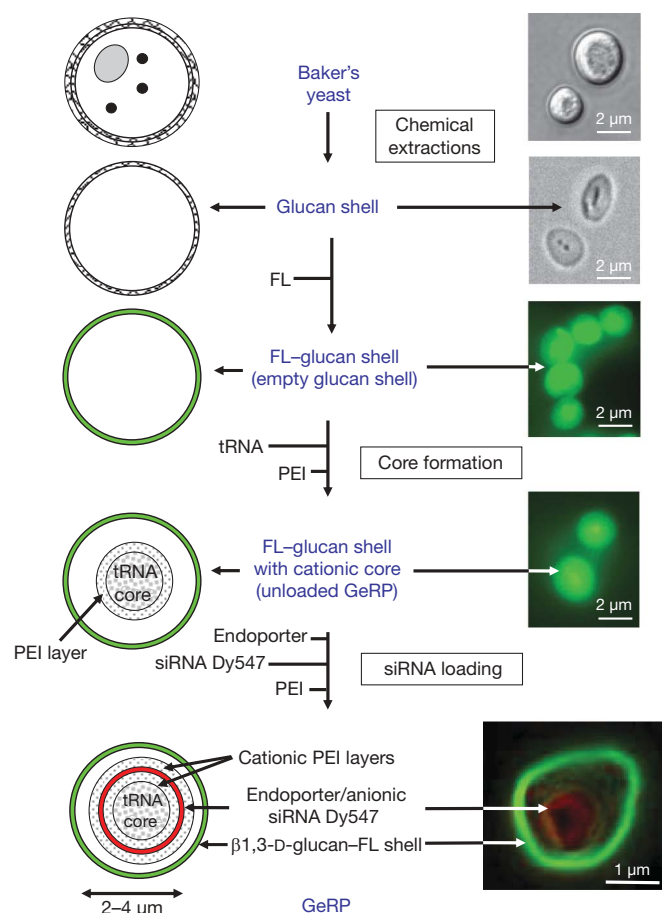


Figure 1 | Production of fluorescent GeRPs. β 1,3-D-glucan particles were purified from baker's yeast by a series of alkaline and solvent extractions, hydrolysing outer cell wall and intracellular components and yielding purified, porous 2–4 μm , hollow β 1,3-D-glucan particles (diagram of particles, left; procedure, middle; microscopy of particles, right). Empty β 1,3-D-glucan particles were then labelled with fluorescein (FL, green). The cores were synthesized by absorbing yeast tRNA, PEI and Dy547-labelled siRNA (red) in a layer-by-layer format. Bottom right: confocal image of a synthesized GeRP.

¹Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

*These authors contributed equally to this work.

Dy547 fluorescent probe and incubated with PECs (10:1 GeRP-to-cell ratio) for 12 h. Confocal microscopy showed that about 90% of the macrophages had internalized at least one fluorescein-GeRP (Fig. 2a, left panels), whereas most cells had internalized multiple fluorescein-GeRPs (Fig. 2a, right panels show two cells at $\times 100$ magnification). Remarkably, a 70–80% knockdown of *Map4k4* mRNA was achieved in 10^6 PECs with as little as 40 pmoles siRNA within GeRPs (Fig. 2b), whereas PBS, unloaded GeRPs or those containing scrambled siRNA had no effect (Fig. 2c).

Lipopolysaccharide (LPS), a major structural component of the outer membrane of Gram-negative bacteria, activates monocytes and macrophages to produce inflammatory cytokines, such as *Tnf- α* and interleukin (IL)-1 β ¹¹. PECs (10^6) were incubated with 10^7 GeRPs containing 40 pmoles of scrambled or *Map4k4* siRNA for 48 h, and then treated with saline or LPS for an additional 6 h. *Tnf* mRNA levels were decreased by 40% in unstimulated cells treated with *Map4k4*-siRNA-containing GeRPs compared to GeRPs containing scrambled siRNA (Fig. 2d). *Map4k4* silencing inhibited the LPS-induced *Tnf* expression by 50% (Fig. 2d) or more (Supplementary Fig. 2b). *Map4k4* silencing in PECs also resulted in an average 30% decrease of LPS-induced *Tnf- α* protein secreted into the medium (Fig. 2e). GeRPs with scrambled siRNA, unloaded GeRPs (GeRPs containing

transfer RNA/PEI cores, but no siRNA) or phosphate buffered saline (PBS) failed to affect *Tnf* expression (Fig. 2c) or secretion (Supplementary Fig. 3), or expression of interferon response genes (Supplementary Fig. 12).

To determine whether depletion of *Map4k4* mRNA in the GeRP-treated macrophages, as shown in Fig. 2b, is mediated by an RNA interference (RNAi)-dependent gene-silencing mechanism, we performed rapid amplification of 5' complementary DNA ends (5' RACE) analysis to identify the cleavage sites of *Map4k4* mRNA¹². After treatment of PECs with GeRPs containing *Map4k4* siRNA, but not scrambled siRNA, a unique 5' RACE product could be detected on agarose gels (Fig. 2f). Sequence analysis of the cloned polymerase chain reaction (PCR) products revealed that 94 out of 100 products were derived from mRNA cleaved at the predicted cut site (ACTA/TGGC; not shown).

cJun N-terminal kinases 1 and 2 (JNK1 (also known as Mapk8) and JNK2 (Mapk9)), extracellular signal-related kinase 1 (Erk1) and 2 (Erk2) (also known as Mapk3 and Mapk1, respectively), p38 (Mapk14) mitogen-activated protein kinase (MAPK) and nuclear factor κ B (NF κ B) pathways regulate *Tnf- α* production in macrophages^{13,14}. Interestingly, we found that *Map4k4* defines a proinflammatory pathway that activates *Tnf- α* expression independently of the JNK1/2, p38 and NF κ B pathways (Fig. 3) or Erk1/2 (not shown). These data demonstrate that *Map4k4* is a new target for suppression of *Tnf- α* expression in LPS-induced macrophage inflammatory responses.

To evaluate GeRPs for oral delivery of siRNA-mediated gene silencing *in vivo*, mice were given Dy547-conjugated scrambled or *Map4k4* siRNA ($20 \mu\text{g kg}^{-1}$) contained within fluorescein-GeRPs (4×10^9 fluorescein-GeRPs per kg or about 10^8 GeRPs in 200 μl per mouse) by oral gavage once daily for 8 consecutive days (Fig. 4a). Staining the PECs recovered from these mice with the macrophage-specific antibody F4/80 (Emr1)–AF633 (magenta) revealed that the fluorescein-GeRPs containing Dy547-siRNA were taken up by GALT macrophages, which then migrate out into peripheral tissues (Fig. 4b). Co-localization of

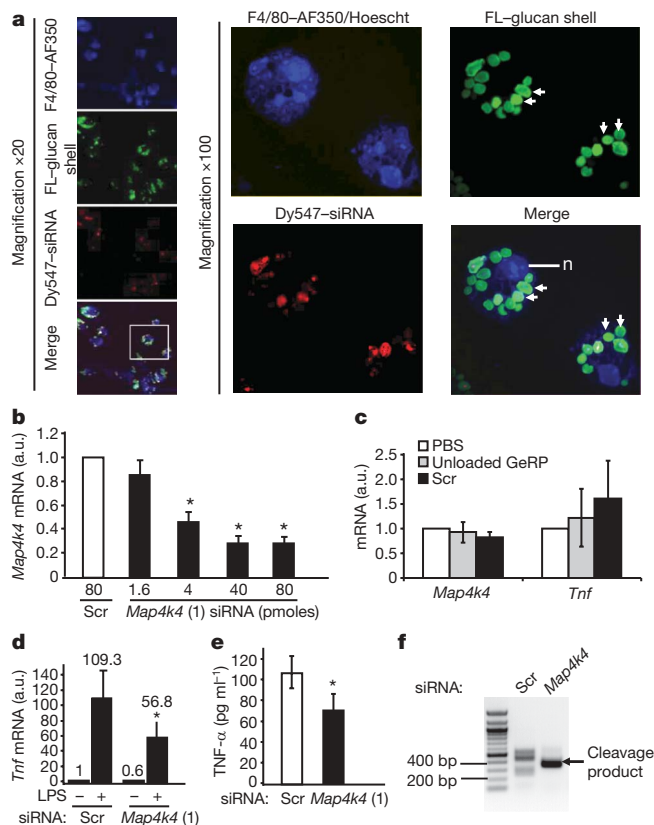


Figure 2 | *In vitro* treatment of GeRPs containing *Map4k4* siRNA silences *Map4k4* expression and inhibits LPS-induced *Tnf- α* production in macrophages. **a**, Confocal images of cultured PECs treated with siRNA-GeRPs (red/green). PECs were stained for F4/80 (blue) and nuclei (n) (Hoescht; blue). Arrows indicate GeRPs. **b**, *Map4k4* mRNA expression from PECs treated with increasing *Map4k4* siRNA concentrations. a.u., arbitrary units. **c**, *Map4k4* and *Tnf* mRNA expression from PECs treated with PBS, unloaded GeRPs or scrambled (Scr) GeRPs. **d**, *Tnf* mRNA expression from PECs treated with Scr- or *Map4k4*-siRNA-containing GeRPs, ± 6 h LPS ($1 \mu\text{g ml}^{-1}$). **e**, *Tnf- α* secretion from PECs treated with Scr or *Map4k4* siRNA-containing GeRPs. Statistical significance was determined by ANOVA and Tukey post test. * $P < 0.05$. Results are mean \pm s.e.m. **f**, PCR amplification of 5' RACE products produces a band (arrow) reflecting the predicted RNAi-mediated mRNA cleavage product in *Map4k4*- but not Scr-siRNA-treated PECs.

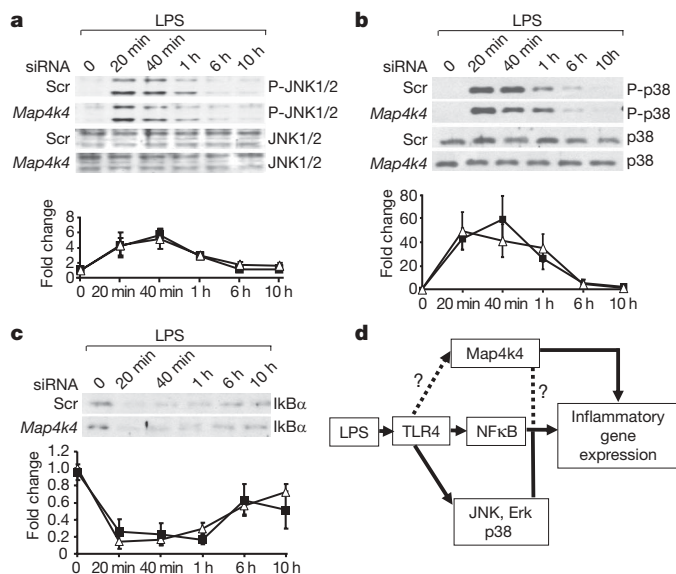


Figure 3 | *Map4k4* silencing fails to affect LPS activation of JNK, p38 and NF κ B signalling pathways. PECs treated with GeRPs containing scrambled (Scr) or *Map4k4* (oligo 1) siRNA, and stimulated with $1 \mu\text{g ml}^{-1}$ LPS for the indicated amounts of time were analysed for phospho- (P-) and total JNK1/2 (**a**) and p38 MAPK (**b**), and for total I κ B α (**c**). Representative blots are shown from three experiments. Immunoblot signals are expressed in arbitrary units (a.u.) and data represent mean \pm s.e.m. ($n = 3$). Filled squares, Scr siRNA; open triangles, *Map4k4* siRNA. **d**, Schematic diagram of potential *Map4k4* signalling to modulate the expression of inflammatory genes such as *Tnf* and *Il1b*.

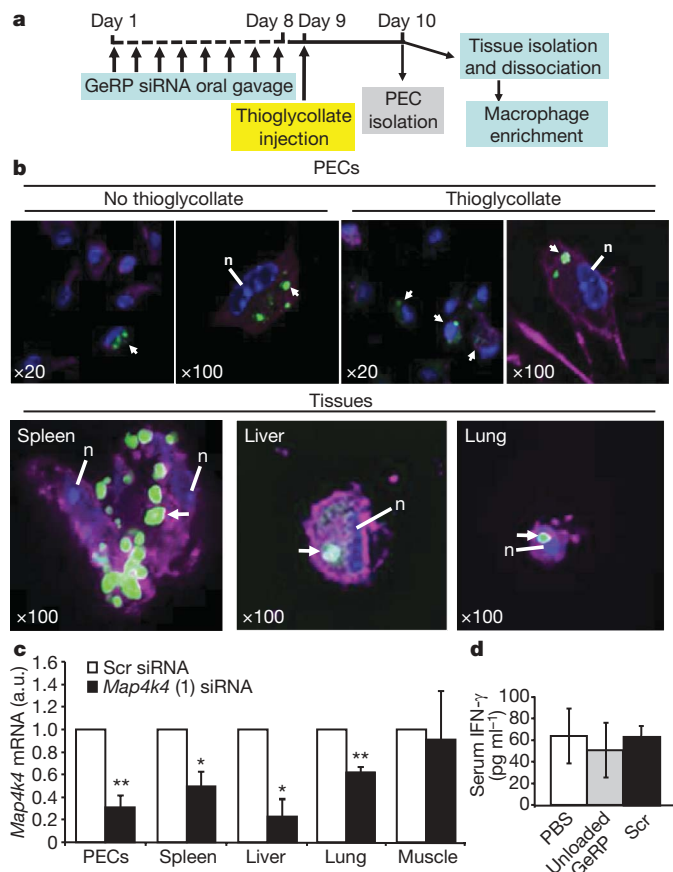


Figure 4 | Orally administered GeRPs containing *Map4k4* siRNA attenuate *Map4k4* mRNA expression in PECs and macrophages from spleen, lung and liver. **a**, Time line of GeRP (scrambled (Scr) or *Map4k4*, oligo 1) administration and PEC/tissue isolation. **b**, Confocal microscopy of PECs and tissue macrophages containing GeRPs (green). Macrophages were stained with F4/80-AF633 (magenta). Arrows indicate cells containing GeRPs. Nuclei (n) were stained with 4,6-diamidino-2-phenylindole (DAPI, blue). Magnification is indicated in the panels. **c**, *Map4k4* mRNA expression in PECs and adherent cells from tissues. **d**, Serum IFN- γ levels from mice gavaged with PBS, unloaded GeRPs or GeRPs containing 20 $\mu\text{g kg}^{-1}$ of Scr siRNA (*n* = 5). Statistical significance was determined by ANOVA and Tukey post test. ***P* < 0.01 and **P* < 0.05. Results are mean \pm s.e.m.

AlexaFluor633, fluorescein-GeRPs and Dy547 fluorescent signals in adherent macrophages was readily observed. Notably, *Map4k4* mRNA expression was inhibited by 70% in PECs isolated from mice orally gavaged with *Map4k4*-siRNA-containing GeRPs compared to scrambled-siRNA-containing GeRPs (Fig. 4c). This level of knockdown is much greater than the relatively lower number of macrophages that could be observed to contain GeRPs in Fig. 4b, indicating possible degradation of the GeRPs and loss of detectable signal even though siRNA-mediated knockdown persists.

Map4k4 silencing was also analysed in macrophages that had migrated to other tissues on day 10 of the protocol. Significant depletions of about 50%, 80% and 40% in *Map4k4* mRNA levels were observed in macrophage-enriched cells isolated from spleen, liver and lung tissues, respectively, in mice treated with *Map4k4*-siRNA-containing GeRPs compared to scrambled-siRNA-containing GeRPs (Fig. 4c). In a parallel experiment, we could identify fluorescein-GeRP-containing macrophages isolated from spleen, liver and lung tissues of mice orally gavaged with fluorescently labelled GeRPs (Fig. 4b, bottom panel). Only a small proportion of macrophages enriched from these tissues contained GeRPs when examined by confocal microscopy (not shown). Tissue sections analysed by fluorescence microscopy also revealed infiltration of spleen, liver and lung

with macrophages containing fluorescein-glucan shells (Supplementary Fig. 4b). Consistent with the lack of gene silencing in cells from skeletal muscle (Fig. 4c), GeRP-containing macrophages were rare in this tissue (Supplementary Fig. 4b). Taken together, these data indicate that macrophages in the GALT internalize orally absorbed GeRPs, undergo siRNA-mediated gene silencing and migrate into tissues throughout the body.

We confirmed that gene silencing by orally delivered GeRPs can be mediated by multiple siRNAs by using a different *Map4k4* siRNA or either of two *Tnf* siRNA oligonucleotides previously found to be effective on macrophages *in vitro* (Supplementary Figs 2a, c and d and 5). We also found efficacy of *Map4k4*-siRNA-containing GeRPs to silence *Map4k4* expression in macrophages after delivery by intraperitoneal (i.p.) injection (Supplementary Fig. 6c). Gene-specific silencing following oral delivery of GeRPs containing siRNA directed against fatty acid binding protein 4, which had no effect on *Map4k4* or *Tnf* expression, was also documented (data not shown). Importantly, oral gavage of GeRPs containing either siRNA or no siRNA (unloaded GeRPs) did not change interferon- γ levels in serum (Fig. 4d), consistent with results *in vitro* (Supplementary Fig. 12). Serum levels of the liver enzymes, aspartate aminotransferase (AST) and alanine aminotransferase (ALT) were within normal ranges (AST < 255 IU l $^{-1}$; ALT < 77 IU l $^{-1}$; Supplementary Fig. 13). Initial experiments indicate that the gene silencing with unmodified siRNA lasted about 8 days after the termination of oral GeRPs administration (not shown). Thus, efficient knockdown of three genes with five different siRNA sequences using orally delivered GeRPs has been achieved (Fig. 4c, Supplementary Fig. 5, and data not shown).

Notably, an 80% decrease in *Tnf* expression in PECs was observed in mice orally gavaged with GeRPs containing *Map4k4* siRNA versus scrambled siRNA (Fig. 5b). This was accompanied by an equally marked 80% knockdown of the inflammatory cytokine Il-1 β , but not Il-10 nor chemokine receptor 2 (*Ccr2*), known to be downregulated by LPS¹⁵. We found no effect of oral gavage of empty β 1,3-D-glucan shells (Supplementary Fig. 9), β 1,3-D-glucan shells containing tRNA and PEI, or GeRPs containing scrambled siRNA on serum *Tnf*- α levels before LPS treatment compared to PBS administration (Supplementary Fig. 11). Importantly, *Tnf* siRNA does not silence Il-1 β , showing specificity of this broad anti-inflammatory response to *Map4k4* knockdown (Supplementary Fig. 7).

Consistent with previous reports¹⁶, we found circulating *Tnf*- α levels are strongly increased 1.5 h after LPS/D-galactosamine (D-GalN) injection into mice and then decreased to basal levels after 4 h. Oral delivery of GeRPs containing *Map4k4* siRNA blocked the increase in serum *Tnf*- α protein (Fig. 5c) and *Tnf*- α levels in peritoneal fluid 1.5 h after LPS/D-GalN injection (Fig. 5d). Similarly, serum *Tnf*- α levels in response to LPS were decreased in mice orally gavaged with GeRPs containing other *Map4k4* or *Tnf* siRNA oligonucleotides (Supplementary Fig. 8a), but not with scrambled siRNA or unloaded GeRPs (Supplementary Fig. 11a). These data demonstrate downregulation of the *Tnf*- α response to an inflammatory stimulus through oral delivery of *Map4k4*-siRNA-containing GeRPs.

The lethality observed in LPS/D-GalN-challenged animals is attributed to inflammatory cytokine toxicity and can be mimicked by administration of *Tnf*- α and Il-1 β , which synergize with each other¹⁷. We therefore tested whether *Map4k4*-siRNA-containing GeRPs could protect against this toxicity. Figure 5e shows that 90% of the control mice orally gavaged with scrambled-siRNA-containing GeRPs before LPS/D-GalN injection died between 4 h and 8 h after LPS/D-GalN injection, whereas 50% of mice treated with *Map4k4*-siRNA-containing GeRPs survived for 8 h and 40% survived the LPS challenge long term (Supplementary Tables 2 and 3). Similarly, protection was obtained with the alternate *Map4k4* siRNA (oligo 2) and two *Tnf* siRNA species (Supplementary Fig. 8b). Hepatocyte apoptosis in response to LPS injection¹⁸ was also attenuated by orally delivered *Map4k4*-siRNA-containing GeRPs (Supplementary Fig. 10), whereas serum insulin and glucose levels were unaffected (Supplementary Fig. 14). Thus, oral

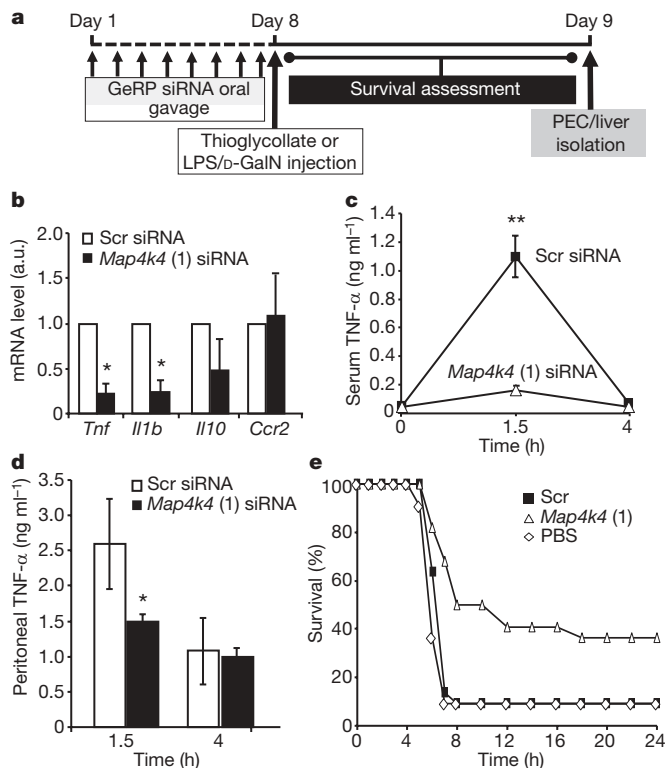


Figure 5 | *Map4k4* silencing by oral gavage with GeRPs inhibits LPS-induced $Tnf-\alpha$ production in vivo. **a**, Time line of siRNA and LPS/D-GalN administration. **b**, *Tnf*, *Il1b*, *Il10* and *Ccr2* mRNA expression in unstimulated PECs from mice gavaged with Scr- and *Map4k4*-siRNA-containing GeRPs ($n = 3$). Statistical significance was determined by a two-tailed Student's *t*-test. **c**, **d**, Serum (**c**) and peritoneal (**d**) $Tnf-\alpha$ levels in siRNA-treated mice 1.5 h and 4 h after LPS/D-GalN administration ($n = 10$). Statistical significance was determined by ANOVA and Tukey post test; ** $P < 0.001$ and * $P < 0.05$. **e**, Survival percentage of mice gavaged with PBS or Scr- and *Map4k4*-siRNA-containing GeRPs, followed by LPS/D-GalN administration (PBS, $n = 11$; Scr and *Map4k4*, $n = 22$). Statistical significance was determined by Kaplan–Meier analysis and Mantel–Cox testing ($P < 0.01$; see Supplementary Table 3). Results are the mean of three independent experiments. All results are mean \pm s.e.m.

gavage of *Map4k4*-siRNA-containing GeRPs significantly protects mice from LPS/D-GalN-induced lethality through inhibition of $Tnf-\alpha$ and $Il-1\beta$ production in macrophages.

The *in vivo* potency of 20 μ g siRNA per kg in GeRPs to mediate gene silencing is 5 to 250 times greater than that in previous studies reporting systemic delivery. Intravenous injection of siRNA formulations requires doses ranging from 125 μ g kg $^{-1}$ to 50 mg kg $^{-1}$ in mice^{19–24} and 1 mg kg $^{-1}$ in nonhuman primates²⁵. For attenuation of LPS-induced lethality in mice by i.p. injection, 1.2 mg *Tnf* siRNA per kg was required²⁶. The high potency of orally delivered siRNA within GeRPs (20 μ g kg $^{-1}$) is even more surprising because unmodified siRNA was used in our studies. This high potency is probably due to protection of siRNA against nuclease degradation by PEI within GeRPs, low nonspecific binding of GeRPs en route to the gut, and high efficiency of GeRP uptake by phagocytic cells in the GALT. Furthermore, the siRNA loading capacity within β 1,3-D-glucan shells is far greater than we used here, and can potentially deliver combinations of siRNA, DNA, proteins and small molecules. We also cannot rule out the possibility that other cell types are also targeted by oral delivery of GeRPs.

Injectable anti- $TNF-\alpha$ protein therapeutics are successful commercial products for the treatment of rheumatoid arthritis, ankylosing spondylitis, Crohn's disease and psoriasis²⁷. Macrophage-mediated pathogenesis is also well characterized in mouse models of obesity-associated insulin resistance²⁸ and atherosclerosis²⁹, whereas such

autoimmune diseases as type 1 diabetes involve the deleterious actions of inflammatory cytokines³⁰. Further development of GeRP-mediated delivery of siRNA to attenuate inflammation for these and other human maladies will be a major focus of our future studies.

METHODS SUMMARY

Preparation of GeRPs. Small interfering RNA was incorporated into the interior of porous, hollow β 1,3-D-glucan shells purified from baker's yeast by a layer-by-layer synthesis strategy to make GeRPs (Fig. 1). In brief, carrier tRNA was absorbed into β 1,3-D-glucan shells and cationic complexes formed by PEI trapping⁷. Negatively charged siRNA was absorbed onto encapsulated positively charged complexes and coated with a PEI layer to produce multi-layered GeRP formulations.

Cell culture and GeRP treatment. PECs were isolated from ten-week-old C57BL/6/J male mice 1–5 days following i.p. injection with thioglycollate broth. PECs were incubated for 48 h with GeRPs or fluorescein–GeRPs at a 10:1 particle-to-cell ratio.

GeRP administration. Ten-week-old C57BL/6/J male mice were injected i.p. daily for 3 days, or gavaged daily for 8 days, with 2×10^9 or 4×10^9 GeRPs per kg (200 μ l), respectively, containing 20 μ g kg $^{-1}$ scrambled or *Map4k4* siRNA.

5' RACE and nested PCR. Messenger RNA was purified from PECs treated for 48 h with GeRPs loaded with scrambled or *Map4k4* siRNA. The GeneRacer Kit (Invitrogen) was used to produce 5' RACE products, followed by nested PCR. GeneRacer primers and gene-specific primers were used. Products were cloned into the pCR4.1 TOPO vector for sequencing.

LPS lethality test. Ten-week-old C57BL/6/J male mice were treated with GeRPs containing scrambled or *Map4k4* siRNA, and then i.p. injected with 25 mg of D-galactosamine and 0.25 μ g *Escherichia coli* LPS. Animals were monitored for survival over 24 h. Blood and peritoneal fluid were collected 1.5 h and 4 h after LPS/D-GalN injection for secreted $Tnf-\alpha$ measurements.

All procedures involving animals were approved by the Institutional Animal Care and Use Committee at University of Massachusetts Medical School.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 November 2008; accepted 6 January 2009.

1. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
2. Grimm, D. & Kay, M. A. Therapeutic application of RNAi: is mRNA targeting finally ready for prime time? *J. Clin. Invest.* **117**, 3633–3641 (2007).
3. Duffield, J. S. The inflammatory macrophage: a story of Jekyll and Hyde. *Clin. Sci. (Lond.)* **104**, 27–38 (2003).
4. Beier, R. & Gebert, A. Kinetics of particle uptake in the domes of Peyer's patches. *Am. J. Physiol.* **275**, G130–G137 (1998).
5. Herre, J., Gordon, S. & Brown, G. D. Dectin-1 and its role in the recognition of β -glucans by macrophages. *Mol. Immunol.* **40**, 869–876 (2004).
6. Vazquez-Torres, A. *et al.* Extraintestinal dissemination of *Salmonella* by CD18-expressing phagocytes. *Nature* **401**, 804–808 (1999).
7. Soto, E. R. & Ostroff, G. R. Characterization of multilayered nanoparticles encapsulated in yeast cell wall particles for DNA delivery. *Bioconjug. Chem.* **19**, 840–848 (2008).
8. Tang, X. *et al.* An RNA interference-based screen identifies MAP4K4/NIK as a negative regulator of PPAR γ , adipogenesis, and insulin-responsive hexose transport. *Proc. Natl Acad. Sci. USA* **103**, 2087–2092 (2006).
9. Tesz, G. J. *et al.* Tumor necrosis factor α (TNF α) stimulates Map4k4 expression through TNF α receptor 1 signaling to c-Jun and activating transcription factor 2. *J. Biol. Chem.* **282**, 19302–19312 (2007).
10. Bouzakri, K. & Zierath, J. R. MAP4K4 gene silencing in human skeletal muscle prevents tumor necrosis factor- α -induced insulin resistance. *J. Biol. Chem.* **282**, 7783–7789 (2007).
11. Lu, Y. C., Yeh, W. C. & Ohashi, P. S. LPS/TLR4 signal transduction pathway. *Cytokine* **42**, 145–151 (2008).
12. Frank-Kamenetsky, M. *et al.* Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proc. Natl Acad. Sci. USA* **105**, 11915–11920 (2008).
13. Barthel, R. *et al.* Regulation of tumor necrosis factor α gene expression by mycobacteria involves the assembly of a unique enhanceosome dependent on the coactivator proteins CBP/p300. *Mol. Cell. Biol.* **23**, 526–533 (2003).
14. Tsytsykova, A. V. *et al.* Post-induction, stimulus-specific regulation of tumor necrosis factor mRNA expression. *J. Biol. Chem.* **282**, 11629–11638 (2007).
15. Zhou, Y., Yang, Y., Warr, G. & Bravo, R. LPS down-regulates the expression of chemokine receptor CCR2 in mice and abolishes macrophage infiltration in acute inflammation. *J. Leukoc. Biol.* **65**, 265–269 (1999).
16. Endo, Y. *et al.* Enhancement by galactosamine of lipopolysaccharide (LPS)-induced tumour necrosis factor production and lethality: its suppression by LPS pretreatment. *Br. J. Pharmacol.* **128**, 5–12 (1999).
17. Okusawa, S., Gelfand, J. A., Ikejima, T., Connolly, R. J. & Dinarello, C. A. Interleukin 1 induces a shock-like state in rabbits. Synergism with tumor necrosis factor and the effect of cyclooxygenase inhibition. *J. Clin. Invest.* **81**, 1162–1172 (1988).

18. Silverstein, R. D-galactosamine lethality model: scope and limitations. *J. Endotoxin Res.* **10**, 147–162 (2004).
19. Filleur, S. *et al.* siRNA-mediated inhibition of vascular endothelial growth factor severely limits tumor resistance to antiangiogenic thrombospondin-1 and slows tumor vascularization and growth. *Cancer Res.* **63**, 3919–3922 (2003).
20. McCaffrey, A. P. *et al.* RNA interference in adult mice. *Nature* **418**, 38–39 (2002).
21. Peer, D., Zhu, P., Carman, C. V., Lieberman, J. & Shimaoka, M. Selective gene silencing in activated leukocytes by targeting siRNAs to the integrin lymphocyte function-associated antigen-1. *Proc. Natl Acad. Sci. USA* **104**, 4095–4100 (2007).
22. Song, E. *et al.* Antibody mediated *in vivo* delivery of small interfering RNAs via cell-surface receptors. *Nature Biotechnol.* **23**, 709–717 (2005).
23. Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432**, 173–178 (2004).
24. Wesche-Soldato, D. E. *et al.* *In vivo* delivery of caspase-8 or Fas siRNA improves the survival of septic mice. *Blood* **106**, 2295–2301 (2005).
25. Zimmermann, T. S. *et al.* RNAi-mediated gene silencing in non-human primates. *Nature* **441**, 111–114 (2006).
26. Sorensen, D. R., Leirdal, M. & Sioud, M. Gene silencing by systemic delivery of synthetic siRNAs in adult mice. *J. Mol. Biol.* **327**, 761–766 (2003).
27. Shealy, D. J. & Visvanathan, S. Anti-TNF antibodies: lessons from the past, roadmap for the future. *Handb. Exp. Pharmacol.* **181**, 101–29 (2008).
28. Ferrante, A. W. Jr. Obesity-induced inflammation: a metabolic dialogue in the language of inflammation. *J. Intern. Med.* **262**, 408–414 (2007).
29. Hansson, G. K. & Libby, P. The immune response in atherosclerosis: a double-edged sword. *Nature Rev. Immunol.* **6**, 508–519 (2006).
30. Shoda, L. K. *et al.* A comprehensive review of interventions in the NOD mouse and implications for translation. *Immunity* **23**, 115–126 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We appreciate critical reading of the manuscript and suggestions by C. Mello, V. Ambros, G. Hannon, S. Corvera and J. Sullivan. We thank P. Zamore for advice on methods. We also appreciate the technical help of P. Furcinitti, A. Burkhart and A. Goller. This work was supported by The University of Massachusetts Diabetes and Endocrinology Center (DK 32520), including its Genomics, Bioinformatics and Imaging Cores, the Diabetes Genome Anatomy Project (DK 60837), Commonwealth Medicine and NIH grant DK 30898.

Author Contributions G.R.O., M.A. and M.P.C. initially conceptualized the study. M.A., G.J.T., M.W., M.C. and S.M.N. performed experiments, and all authors participated in designing experiments, and analysing and interpreting data. M.A. and G.J.T. contributed equally to this work. G.R.O. and E.S. provided the β 1,3-D-glucan-shell-encapsulated cationic materials and they with M.A. and G.J.T. developed the GeRP formulations used in these studies. M.A., G.R.O. and M.P.C. wrote the manuscript.

Author Information The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.P.C. (michael.czech@umassmed.edu) and G.R.O. (gary.ostroff@umassmed.edu).

METHODS

Preparation of β 1,3-D-glucan shells. β 1,3-D-glucan shells were prepared as described previously⁷. Baker's yeast (AB Mauri Food Inc.) was alkaline-and-acid-extracted and dried at 20–25 °C following alcohol and acetone extractions. β 1,3-D-glucan shells are 2–4 μ m hollow, porous microspheres consisting primarily of β 1,3-D-glucan and typically contain 5×10^{11} particles per g⁷.

Fluorescein labelling of β 1,3-D-glucan shells. β 1,3-D-glucan shells (1 g) were washed with sodium carbonate buffer (0.1 M, pH 9.2) and resuspended in 0.1 l carbonate buffer. 5-(4,6-dichlorotriazinyl) amino fluorescein (Invitrogen; 1 mg ml⁻¹ in DMSO) was added to the buffered β 1,3-D-glucan shell suspension (10% v/v) and mixed at 20–25 °C in the dark for 16 h. Tris buffer (2 mM) was added, incubated for 15 min and β 1,3-D-glucan shells washed with sterile pyrogen-free water until the colour was removed. The β 1,3-D-glucan shells were then dehydrated with absolute ethanol and acetone and dried in the dark at 20–25 °C.

Preparation of GeRPs. For empty GeRPs, dry β 1,3-D-glucan shells were mixed with a volume of yeast RNA (Sigma; 10 mg ml⁻¹ in 50 mM Tris HCl, pH 8, 2 mM EDTA and 0.15 M NaCl (TEN)) to minimally swell the shells, and were incubated for 2 h at room temperature to allow for RNA absorption as described previously⁷. Neutral PEI (Aldrich; 25 kDa branched PEI; 2 mg ml⁻¹ in TEN, pH 7) was added in excess to form β 1,3-D-glucan -shell-encapsulated RNA complexes, and these were resuspended by homogenization and incubated for at least 1 h. The β 1,3-D-glucan-shell-encapsulated cationic complexes were sterilized in 70% ethanol, washed and resuspended in sterile pyrogen-free saline, counted by a haemocytometer (200 \times), diluted to 1×10^9 shells per ml in saline and stored at -20 °C.

For loading of siRNA into GeRPs for *in vitro* experiments, 1×10^7 empty GeRPs were diluted with sterile saline and incubated with 1 nmole Endo-Porter (Gene Tools) for 1 h at 20–25 °C. Next, 40 pmoles of siRNA was added and incubated at 20–25 °C for 2 h (all of the *in vitro* experiments contained 40 pmoles siRNA with the exception of the *in vitro* dose response, which loaded 1.6, 4, 40 or 80 pmoles of siRNA). Then, PEI (5 μ g in saline) was added while vortexing and incubated at 20–25 °C for 20 min to trap the siRNA in the GeRPs. The PEI was quenched by the addition of 600 μ l of complete DMEM. siRNA-loaded GeRPs were added to cells in culture at a concentration of 1×10^7 GeRPs per 1×10^6 cells. For loading of siRNA into GeRPs for *in vivo* experiments, the same procedure was followed with an adjusted ratio of 40 pmoles of siRNA in 1×10^8 empty GeRPs per dose. Empty GeRPs were incubated with 1 nmole Endo-Porter for 1 h at 20–25 °C. Next, 40 pmoles of siRNA was added, incubated at 20–25 °C for 2 h and trapped with PEI (50 μ g) for 20 min at 20–25 °C. The siRNA-loaded GeRPs were then washed, resuspended in saline and sonicated to ensure homogeneity of the GeRP preparation. For each experiment, GeRP batches were aliquoted into tubes for daily dosing, flash-frozen in liquid nitrogen and stored at -20 °C. The final oral dose of GeRPs in 200 μ l contains 40 pmole siRNA in 1×10^8 GeRPs.

Tissue macrophage isolation. C57BL/6J male mice (ten-week-old) were administered 1×10^8 GeRPs by daily oral gavage for 8 days. On day 9, mice were i.p. injected with thioglycollate. Cells from spleen, liver, lung, muscle and PECs were isolated on day 10. Tissues were cut into small pieces, washed with PBS and digested at 37 °C for 30 min with agitation using 5 mg ml⁻¹ collagenase. Digested tissues were filtered through 70- μ m pore nylon mesh, centrifuged, and the cells were plated in plastic dishes for 2–3 h in medium (DMEM plus 10% fetal bovine serum). The cells were washed with PBS to remove non-adherent cells, and adherent cells were used for real-time PCR and microscopy.

For confocal experiments, mice were gavaged with a single 2×10^8 dose of GeRPs, and tissue macrophages isolated after 24 h as described previously.

Microscopy. Fixed cells from *in vitro* and *in vivo* experiments were incubated with F4/80 primary antibody (AbD Serotec) followed by AlexaFluor350 or 633 secondary antibody (Invitrogen). Nuclei were stained with Hoechst 33342 or DAPI as denoted in the figure legends.

Isolation of RNA and real-time PCR. RNA isolation was performed according to the Trizol reagent protocol (Invitrogen). Complementary DNA was synthesized from 1 μ g of total RNA using iScript cDNA synthesis kit (Bio-Rad) according to the manufacturer's instructions. For real-time PCR, synthesized cDNA forward and reverse primers along with the iQ SYBR Green Supermix were run on the MyIQ real-time PCR system (Bio-Rad). Sequences of the primers used were designed with Primer Bank (Supplementary Table 4). The ribosomal mRNA 36B4 was used as an internal loading control, as its expression did not change over a 24 h period with the addition of LPS, TNF- α or siRNA against the genes used in this study.

ELISA assay. Tnf- α levels in the PEC supernatant, plasma and peritoneal fluids and interferon- γ levels in plasma were measured using mouse ELISA kits (Pierce) as recommended by the manufacturer.

AST and ALT measurement. To test for liver toxicity, levels of ALT and AST activity in serum were measured using a commercial kit (Fisher Scientific) according to the manufacturer's instructions.

Histology and TUNEL assay. Tissue sections were stained with F4/80–AlexaFluor405 antibody (AbD-Serotec) and were haematoxylin-stained. TUNEL assay was performed on liver sections from mice challenged with LPS/D-GalN according to the manufacturer's instructions (Upstate). TUNEL images were obtained using a Zeiss Axiovert 200 inverted microscope equipped with a Zeiss AxioCam HR CCD camera with 1,300 \times 1,030 pixels basic resolution and a Zeiss Plan NeoFluar 20 \times /0.50 Ph2 (DIC II) objective. Zeiss Aplanachromat 100 \times /1.40 Oil (440780).

Tissue and cell images were obtained with a Solamere CSU10 Spinning Disk confocal system mounted on a Nikon TE2000-E2 inverted microscope. Images were taken with a multi-immersion \times 20 objective with a NA = 0.75, oil:W.D. = 0.35 mm, or a \times 100 Plan Apo VC objective NA = 1.4, oil:W.D. = 0.13 mm.

Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay

Kazufumi Matsushita^{1,3*}, Osamu Takeuchi^{1,3*}, Daron M. Standley², Yutaro Kumagai^{1,3}, Tatsukata Kawagoe^{1,3}, Tohru Miyake^{1,3}, Takashi Satoh^{1,3}, Hiroki Kato^{1,3}, Tohru Tsujimura⁴, Haruki Nakamura⁵ & Shizuo Akira^{1,3}

Toll-like receptors (TLRs) recognize microbial components, and evoke inflammation and immune responses^{1–3}. TLR stimulation activates complex gene expression networks that regulate the magnitude and duration of the immune reaction. Here we identify the TLR-inducible gene *Zc3h12a* as an immune response modifier that has an essential role in preventing immune disorders. *Zc3h12a*-deficient mice suffered from severe anaemia, and most died within 12 weeks. *Zc3h12a*^{−/−} mice also showed augmented serum immunoglobulin levels and autoantibody production, together with a greatly increased number of plasma cells, as well as infiltration of plasma cells to the lung. Most *Zc3h12a*^{−/−} splenic T cells showed effector/memory characteristics and produced interferon- γ in response to T-cell receptor stimulation. Macrophages from *Zc3h12a*^{−/−} mice showed highly increased production of interleukin (IL)-6 and IL-12p40 (also known as IL12b), but not TNF, in response to TLR ligands. Although the activation of TLR signalling pathways was normal, *Il6* messenger RNA decay was severely impaired in *Zc3h12a*^{−/−} macrophages. Overexpression of *Zc3h12a* accelerated *Il6* mRNA degradation via its 3'-untranslated region (UTR), and destabilized RNAs with 3'-UTRs for genes including *Il6*, *Il12p40* and the calcitonin receptor gene *Calcr*. *Zc3h12a* contains a putative amino-terminal nuclease domain, and the expressed protein had RNase activity, consistent with a role in the decay of *Il6* mRNA. Together, these results indicate that *Zc3h12a* is an essential RNase that prevents immune disorders by directly controlling the stability of a set of inflammatory genes.

The innate immune responses induced by TLRs are tightly controlled, because aberrant activation of TLR responses is harmful to the host, resulting in inflammatory diseases^{1–3}. TLR signalling induces the expression of several genes, although only some of these genes have been functionally characterized as immune response modifiers. Therefore, investigation of TLR-inducible genes is important for clarifying the control mechanisms of innate immune reactions. To examine TLR-induced gene expression comprehensively, we performed microarray analysis using mouse macrophages from wild-type, *Myd88*^{−/−} and *Trif*^{−/−} (also known as *Ticam1*^{−/−}) mice stimulated with lipopolysaccharide (LPS), a TLR4 ligand. We selected 214 genes in which the expression was induced more than twofold either at 1 or 4 h after stimulation in wild-type cells. Hierarchical clustering of these LPS-inducible genes showed that they could be classified into three major clusters (Supplementary Fig. 1a). Among the clusters, genes in cluster III were rapidly induced in a MyD88-dependent manner. This cluster contained, among others, *Tnf*, *Nfkbiz* and *Zfp36*. Cluster III also contained the gene encoding *Zc3h12a* (Supplementary Fig. 1b). Northern blot analysis confirmed that *Zc3h12a* mRNA was rapidly induced in mouse macrophages after LPS stimulation and gradually decreased with time

(Supplementary Fig. 1c). *Zc3h12a* has a CCH-type zinc-finger motif, and forms a family with the homologous proteins *Zc3h12b*, *Zc3h12c* and *Zc3h12d*. Fractionation experiments showed that the *Zc3h12a* protein is mainly localized in the cytoplasm, rather than in the nucleus (Supplementary Fig. 1d).

To investigate the functional roles of *Zc3h12a* in the control of immune responses *in vivo*, we generated *Zc3h12a*-deficient mice (Supplementary Fig. 2a and 2b). PCR with reverse transcription (RT-PCR) analysis confirmed that the expression of *Zc3h12a* was abrogated in *Zc3h12a*^{−/−} macrophages (Supplementary Fig. 2c). Although *Zc3h12a*^{−/−} mice are born in a Mendelian ratio, they showed growth retardation, and most of the mice spontaneously died within 12 weeks of birth (Fig. 1a). *Zc3h12a*^{−/−} mice showed severe splenomegaly and lymphadenopathy (Fig. 1b). Histological examination revealed infiltration of plasma cells in the lung, paraepithelium of the bile duct and pancreas (Fig. 1c and Supplementary Fig. 3). Plasma cells also accumulated in *Zc3h12a*^{−/−} lymph nodes and spleens (Fig. 1c). In the lymph nodes, granuloma formation was observed leading to the generation of giant cells with fused macrophages. Nevertheless, inflammatory changes were not observed in either the intestine or the joints of *Zc3h12a*^{−/−} mice (data not shown).

Zc3h12a^{−/−} mice suffered from severe anaemia, together with an increase in white blood cells and platelets (Fig. 1d). Furthermore, *Zc3h12a*^{−/−} mice developed hyperimmunoglobulinemia of all immunoglobulin isotypes tested (Fig. 1e), and plasma cells infiltrated in the lung interstitial tissues were readily stained with anti-IgG or anti-IgA antibodies (Fig. 1g). Production of anti-nuclear antibodies and anti-double-stranded-DNA antibodies were detected in *Zc3h12a*^{−/−} mice (Fig. 1f). Flow cytometric analysis showed that about 70% of CD19⁺ B cells were IgM[−] IgD[−], but immunoglobulin⁺, indicating that most *Zc3h12a*^{−/−} B cells underwent a class switch in the spleen (Fig. 2a and data not shown). Furthermore, CD138⁺ CD19^{du} plasma cells were abundant in the spleen of *Zc3h12a*^{−/−} mice (Fig. 2b). In addition, the expression of CD69 was upregulated in splenic CD3⁺ T cells, and CD44^{high} CD62L[−] T cells accumulated in the periphery (Fig. 2c and Supplementary Fig. 4a). Nevertheless, the proportion of CD4⁺ Foxp3⁺ regulatory T cells was comparable between wild-type and *Zc3h12a*^{−/−} mice (Supplementary Fig. 4b). Stimulation of splenic T cells with anti-CD3 antibody resulted in increased production of IFN- γ , but not IL-17 (Fig. 2d and Supplementary Fig. 4c). Ter119⁺ (also known as Ly76⁺) erythroblast population was higher in *Zc3h12a*^{−/−} spleens, probably reflecting the responses to anaemia (Supplementary Fig. 4d). However, the ratios of B to T cells and of CD4⁺ to CD8⁺ cells were not altered in *Zc3h12a*^{−/−} spleens (Supplementary Fig. 4e, f). To examine whether haematopoietic cells are sufficient for the development of disease, we transferred bone marrow cells from *Zc3h12a*^{−/−} mice to recipient

¹Laboratory of Host Defense, ²Laboratory of Systems Immunology, WPI Immunology Frontier Research Center, ³Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁴Department of Pathology, Hyogo College of Medicine, 1-1 Mukogawa-cho, Nishinomiya, Hyogo 663-8501, Japan. ⁵Research Center for Structural and Functional Proteomics, Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita, Osaka 565-0871, Japan.

*These authors contributed equally to this work.

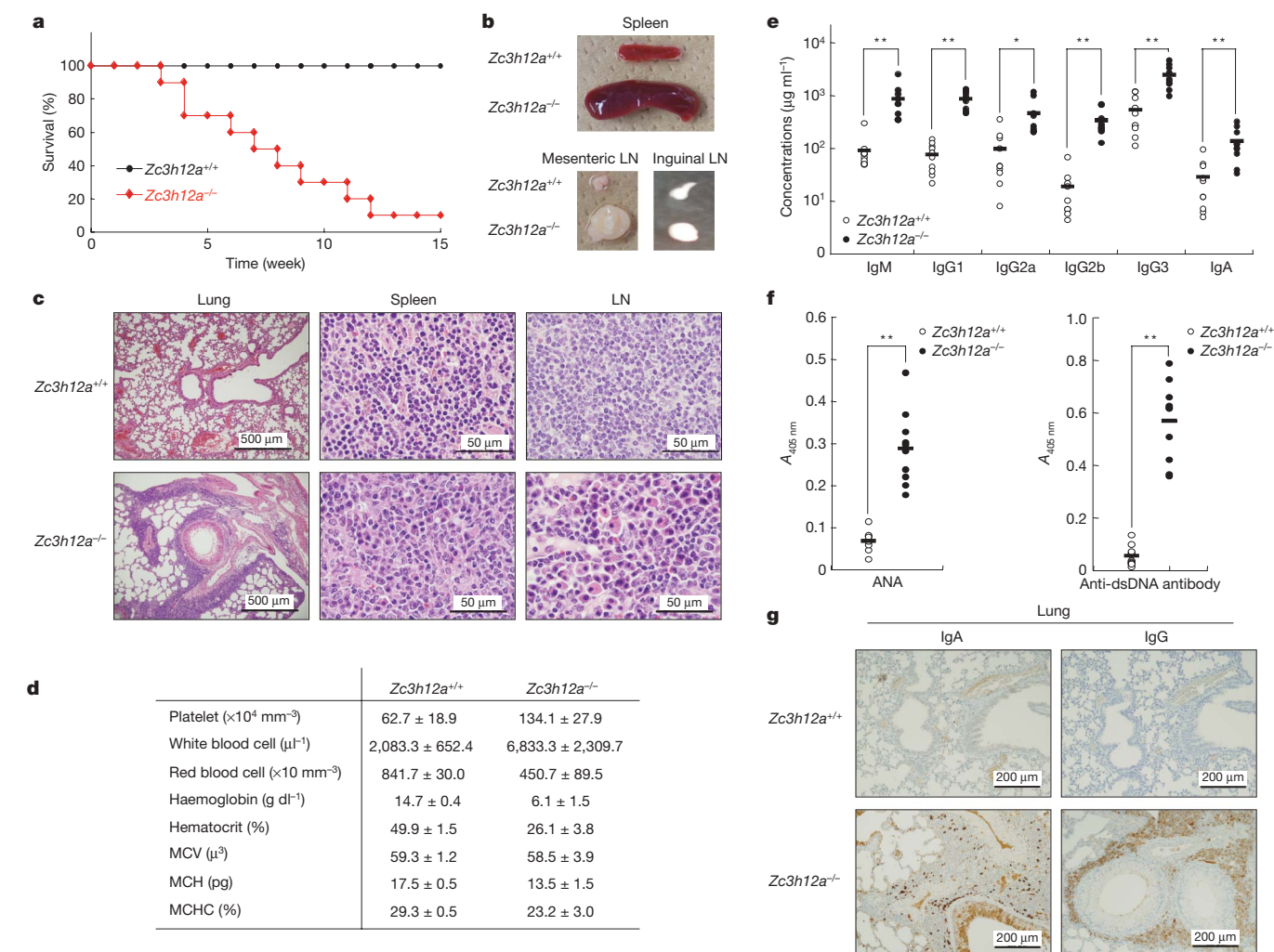


Figure 1 | Early onset of fetal autoimmune disease in $Zc3h12a^{-/-}$ mice. **a**, Survival rates of wild-type ($Zc3h12a^{+/+}$) and $Zc3h12a^{-/-}$ mice at indicated time periods ($n = 10$). **b**, Gross appearance of spleens and mesenteric and inguinal lymph nodes (LN) from wild-type and $Zc3h12a^{-/-}$ mice. **c**, Histology of lung, spleen and lymph nodes from wild-type and $Zc3h12a^{-/-}$ mice. **d**, Examination of blood cells. Data are mean \pm s.d. of six

samples. **e**, Hypergammaglobulinemia in $Zc3h12a^{-/-}$ mice. Serum immunoglobulin levels are shown. **f**, Production of anti-nuclear antibodies (ANA) and anti-double-stranded DNA (anti-dsDNA) antibodies in $Zc3h12a^{-/-}$ mice. Statistical significance in **e** and **f** was determined using the Student's *t*-test. **g**, Immunohistochemistry of lung sections stained with anti-IgG and anti-IgA antibodies.

C57BL/6 mice. $Zc3h12a^{-/-}$ bone marrow chimaeras showed delayed, but marked, development of lymphadenopathy and accumulation of plasma cells and $\text{CD44}^{\text{high}}\text{CD62L}^{-}$ T cells, indicating that haematopoietic cells contribute to the development of immune disorders (Supplementary Fig. 5).

Taken together, these results demonstrate that $Zc3h12a$ is essential for preventing the development of severe immune diseases characterized by an increase in immunoglobulin-producing plasma cells and the formation of granulomas.

We then examined cytokine production from macrophages. As shown in Fig. 2e, stimulation with TLR ligands, MALP-2 (TLR2), poly(I:C) (TLR3), LPS (TLR4), R-848 (TLR7) and CpG-DNA (TLR9), induced increased production of IL-6 and IL-12p40, but not of TNF, in $Zc3h12a^{-/-}$ macrophages. Northern blot analysis showed that *Il6* mRNA, but not *Tnf*, *Cxcl1* or *Nfkb* mRNA, increased significantly in response to LPS in $Zc3h12a^{-/-}$ macrophages, (Fig. 2f). We then performed microarray analysis to assess the difference in LPS-inducible gene expression in wild-type and $Zc3h12a^{-/-}$ macrophages. Microarray analysis of LPS-inducible genes in macrophages showed that most LPS-inducible genes were comparably expressed in wild-type and $Zc3h12a^{-/-}$ cells (Supplementary Fig. 6). Nevertheless, a particular set of genes was highly expressed in $Zc3h12a^{-/-}$ macrophages. These included *Il6*,

Ifng, *Calcr* and *Sprr2d* (Fig. 2g). No differences were observed in the activation of NF- κ B or the activator protein 1 (AP-1) by LPS between wild-type and $Zc3h12a^{-/-}$ macrophages, indicating that $Zc3h12a$ is not involved in the regulation of the initial TLR signalling pathways (Supplementary Fig. 7).

CCCH-type zinc-finger proteins have been implicated in mRNA metabolism such as mRNA splicing, polyadenylation and the regulation of mRNA decay⁴⁻⁶. Thus, we proposed that $Zc3h12a$ might be involved in the destabilization of mRNA, and we examined this possibility using *Il6* as an example. Wild-type and $Zc3h12a^{-/-}$ macrophages were stimulated with LPS for 2 h followed by actinomycin D treatment. The half-life of *Il6* mRNA, but not of *Tnf* or *Cxcl1* mRNA, increased in $Zc3h12a^{-/-}$ macrophages compared to wild-type cells (Fig. 3a, b). These results indicate that $Zc3h12a$ regulates *Il6* mRNA post-transcriptionally. To determine whether $Zc3h12a$ expression controls *Il6* mRNA, we transfected HEK293 cells stably expressing the tetracycline repressor protein fused to the transactivation domain of the viral transcription factor VP-16 (Tet-off 293 cells), with a plasmid containing the *Il6* coding sequence (CDS) with the 3'-UTR sequence under the control of a tetracycline-responsive promoter (TRE) (pTREtight-*Il6*-CDS + 3'-UTR). Treatment with doxycycline (dox) terminated the transcription of *Il6* mRNA, and the mRNA decayed in an incubation time-dependent manner (Fig. 3c).

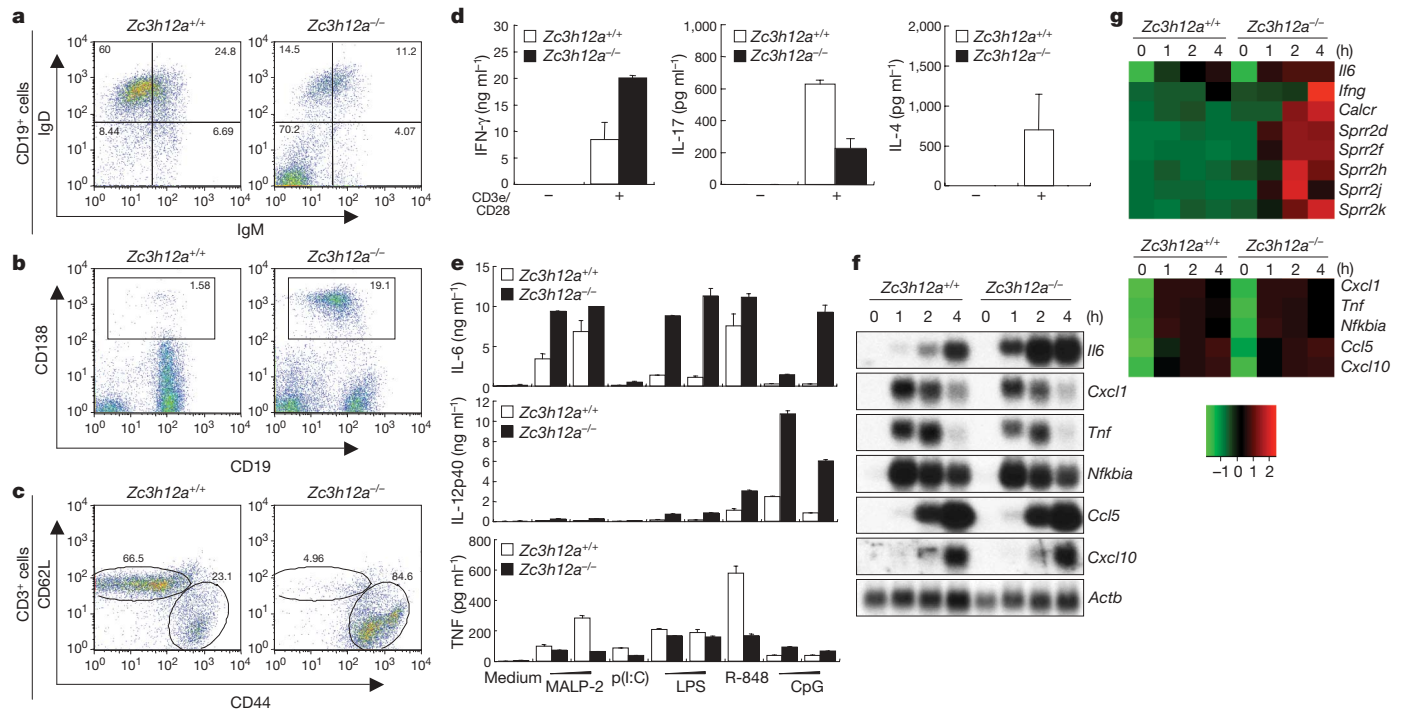


Figure 2 | Cellular abnormalities and augmented cytokine production in *Zc3h12a*^{-/-} mice. **a–c**, Flow cytometric analysis of splenocytes. Expression of IgM and IgD in splenic CD19⁺ B cells (**a**), the proportion of plasma cells (CD138, also known as Sdc1, and CD19) in the spleen (**b**), and expression of CD62L (also known as Sell) and CD44 in splenic T cells (**c**). Similar results were obtained in three independent experiments. **d**, Production of IFN- γ , IL-17 and IL-4 in response to CD3e and CD28 stimulation in splenic T cells. Error bars indicate s.d. of duplicates. Similar results were obtained in three independent experiments. **e**, Peritoneal macrophages from wild-type and *Zc3h12a*^{-/-} mice were stimulated with MALP-2 (1, 10 ng ml⁻¹), poly(I:C)

(p(I:C), 100 μ g ml⁻¹), LPS (10, 100 ng ml⁻¹), R-848 (10 nM) and CpG-DNA (0.1, 1 μ M) for 24 h. The concentrations of IL-6, IL-12p40 and TNF in the culture supernatants were measured by ELISAs. Error bars indicate the s.d. of duplicates. Similar results were obtained in three independent experiments. **f**, Total RNA from macrophages stimulated with LPS (100 ng ml⁻¹) for indicated periods was extracted and subjected to northern blotting for the expression of *Il6*, *Cxcl1*, *Tnf*, *Nfkb1a*, *Ccl5*, *Cxcl10* and β -actin (*Actb*). **g**, Heat map representation of the expression of selected LPS-inducible genes on the basis of microarray analysis of wild-type and *Zc3h12a*^{-/-} peritoneal macrophages.

Overexpression of *Zc3h12a* greatly accelerated the degradation of *Il6* mRNA (Fig. 3c, d). In contrast, *Zc3h12a* did not affect the expression of mRNA harbouring the *Il6* CDS without the 3'-UTR sequence (pTREtight-*Il6*-CDS) (Fig. 3c, d).

Mouse *Il6* mRNA contains five adenine-uridine-rich elements (AREs) in its 3'-UTR (Fig. 3e)⁷. In addition, a conserved element between species comprising about 30 nucleotides was reported to be important for *Il6* mRNA destabilization⁸. To investigate regions of the *Il6* 3'-UTR that are critical for conferring *Zc3h12a* responsiveness, we used a series of luciferase reporter constructs (pGL3) containing several regions of the *Il6* 3'-UTR (Fig. 3e). When full-length *Il6* 3'-UTR (1–403) was inserted into the reporter, the luciferase activity decreased compared to the luciferase reporter alone. Co-expression of *Zc3h12a* further reduced the luciferase activity of pGL3-*Il6* 3'-UTR (1–403) (Fig. 3f). Whereas the luciferase activities of pGL3-*Il6* 3'-UTR (1–70) and pGL3-*Il6* 3'-UTR (172–403) were not altered by the expression of *Zc3h12a*, the luciferase activity of pGL3-*Il6* 3'-UTR (56–173) decreased in the presence of *Zc3h12a*. *Il6* 3'-UTR (56–173) contains two AREs and the conserved element (Fig. 3f). By using a set of luciferase reporter constructs with shortened *Il6* 3'-UTRs, we found that the conserved element, but not the ARE, of *Il6* 3'-UTR was important for destabilization by *Zc3h12a* (Fig. 3f and Supplementary Fig. 8). Although the luciferase activity of pGL3- β -globin 3'-UTR was not affected by *Zc3h12a* expression, addition of the conserved element of *Il6* 3'-UTR (77–108) to β -globin 3'-UTR conferred a response to *Zc3h12a* (Fig. 3g). The expression of *Zc3h12a* reduced the luciferase activity of reporters with the 3'-UTR for *Il12p40* or calcitonin receptor (*Calcr*), but not those with the 3'-UTR of *Ifng* (Fig. 3h), indicating that *Il6*, *Il12p40* and *Calcr* mRNAs are directly regulated by *Zc3h12a*. IFN- γ might be secondarily regulated by the overproduction of IL-12.

We next examined whether *Zc3h12a* directly associates with RNA. Synthesized *Zc3h12a* protein, but not bovine serum albumin (BSA), associates with *Il6* 3'-UTR (1–403) RNA transcribed *in vitro*, indicating that *Zc3h12a* harbours an RNA-binding capacity (Fig. 4a).

Furthermore, we tested whether the CCCH sequence of *Zc3h12a* is critical for its role in *Il6* mRNA decay. The expression of *Zc3h12a* containing the C306R mutation in the CCCH zinc-finger domain, and *Zc3h12a* without the CCCH domain (lacking amino acids 306–322), could still destabilize *Il6* mRNA (Fig. 4b, c), although these mutant proteins had a reduced ability to degrade *Il6* mRNA when low amounts of proteins were expressed (Supplementary Fig. 9). These results indicate that the CCCH motif plays a part in the control of *Il6* mRNA decay.

The above result prompted us to look for another domain(s) responsible for mRNA decay. Sequence alignment indicated that a conserved N-terminal domain (139–297) in *Zc3h12a*, just preceding the zinc-finger domain (300–324), shares remote homology to the PilT N-terminus (PIN) domain-like Structural Classification of Proteins (SCOP) superfamily (Fig. 4d). Structural modelling, followed by alignment to other PIN domain structures, revealed a conserved, negatively charged pocket—formed by Asp 141, Asn 144, Asp 226, Asp 244 and Asp 248—that is potentially important for magnesium binding and enzymatic activity (Fig. 4d, e). We proposed that the N-terminal domain of the *Zc3h12a* protein might be an RNase, and synthesized *Zc3h12a* protein showed RNase activity for *Il6* 3'-UTR (1–403) mRNA in an Mg²⁺-dependent manner (Fig. 4f, g). *Zc3h12a* degraded 5'- and 3'-labelled RNA with similar kinetics, suggesting that *Zc3h12a* has endonuclease activity (Supplementary Fig. 10). The activity of *Zc3h12a* seemed to be largely sequence-independent *in vitro*, because target RNAs with various sequences were degraded almost completely (data not shown). Furthermore, the *Zc3h12a*(D141N) mutant did not

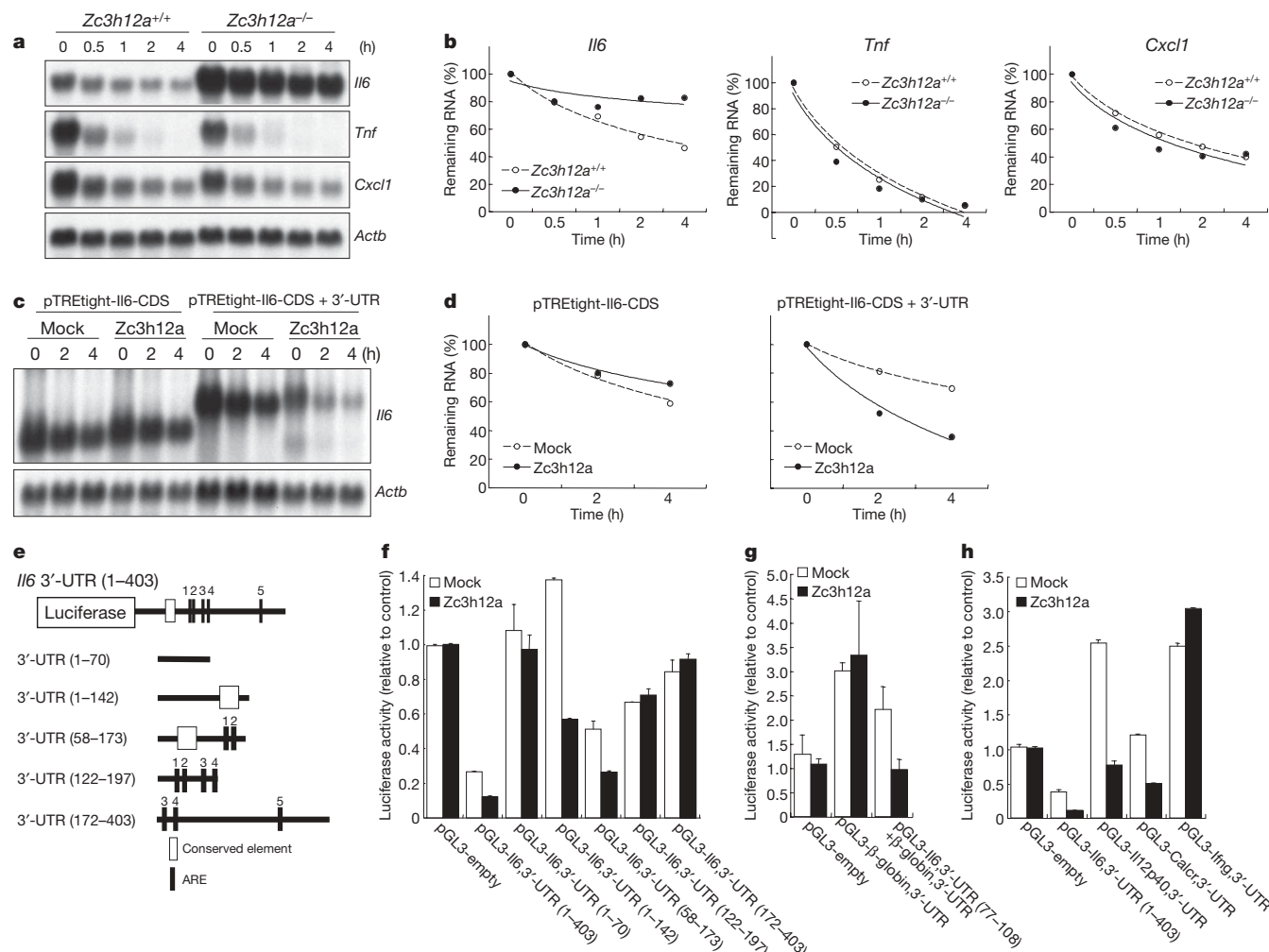


Figure 3 | Zc3h12a destabilizes mRNA from a set of genes through their 3'-UTRs. **a**, **b**, Peritoneal macrophages were treated with LPS (100 ng ml⁻¹) for 120 min and then treated with actinomycin D (2 µg ml⁻¹) for the indicated times. Total RNA (10 µg) was extracted and subjected to RNA blot analysis for the expression of *Il6*, *Tnf*, *Cxcl1* and β -actin (*Actb*) probes (**a**). Similar results were obtained in three independent experiments. The autoradiograph was quantified and the ratio of *Il6*, *Tnf* and *Cxcl1* to *Actb* was used to determine the remaining mRNA levels (**b**). **c**, **d**, HEK293 Tet-off cells were cotransfected with pTREtight-*Il6*-CDS or pTREtight-*Il6*-CDS + 3'-UTR, together with the Zc3h12a expression plasmid or control (empty) plasmid. Cells were divided 3 h after transfection and incubated overnight. Total RNA was prepared after dox (1 µg ml⁻¹) treatment, and *Il6* and *Actb* levels were determined by

degrade RNA, indicating that the conserved pocket indeed functions as an RNase active site (Fig. 4f, h). The Zc3h12a(D141N) mutant failed to destabilize RNA containing the *Il6* 3'-UTR, indicating that the RNase activity is essential for the function of Zc3h12a (Fig. 4c, i).

This study clearly demonstrates that Zc3h12a is essential for the inhibition of the development of severe autoimmune responses culminating in the lethality of mice. Production of IL-6 and IL-12p40, but not TNF, was increased in Zc3h12a^{-/-} macrophages due to mRNA decay failure. CCCH-type zinc-finger proteins have been shown to control mRNA decay by associating with the 3'-UTR. For example, tristetraprolin (Ttp, also known as Zfp36) and its homologues Zfp3611, Zfp3612 and Zfp3613, are critical for the decay of the mRNAs for TNF, GM-CSF, CXCL1 and so on^{4,9}. Aged *Ttp*^{-/-} mice develop autoimmune arthritis owing to TNF production¹⁰. However, to our knowledge, there is no report showing that *Ttp*^{-/-} cells produce increased amounts of IL-6 in response to TLR stimulation. Interestingly, the loss of Zc3h12a did not affect the expression of *Tnf* mRNA in macrophages, indicating that Ttp and Zc3h12a control mRNA decay for different cytokines. Zc3h12a

northern blot analysis (**c**). The autoradiograph was quantified and the ratio of *Il6* to *Actb* was used to determine remaining mRNA levels

(**d**). **e**–**g**, Determination of Zc3h12a responsive regions in the *Il6* 3'-UTR. Schematic presentation of *Il6* 3'-UTR and its deletion constructs (**e**). HEK293 cells were transfected with the indicated pGL3 plasmids containing various sequences of *Il6* 3'-UTR (**f**) and β -globin 3'-UTR (**g**), together with the Zc3h12a expression plasmid or control (empty) plasmid. The luciferase activity was determined after 48 h. **h**, HEK293 cells were transfected with pGL3 harbouring the 3'-UTR for *Il6*, *Il12p40*, *Calcr* or *Ifng*, together with the Zc3h12a expression plasmid or control (empty) plasmid and the luciferase activity was determined after 48 h. Error bars indicate the s.d. of duplicates. Similar results were obtained in three independent experiments.

targeted RNA sequences other than AREs, and the IL-6 AREs seem to be regulated by an unknown Zc3h12a-independent mechanism. Considering the profound pathological findings observed in Zc3h12a^{-/-} mice, genes other than *Il6* and *Il12p40* are probably critically involved in the pathogenesis too. Identification of Zc3h12a target genes in response to other stimuli or in other cell types will improve our knowledge of the whole mechanism of abnormalities observed in Zc3h12a^{-/-} mice. Zc3h12a was recently reported to be a monocyte chemotactic protein-1 (MCP-1)-induced protein¹¹, and overexpression of Zc3h12a protein was shown to suppress cytokine production in macrophages through inhibition of NF- κ B activation¹². However, the present study shows that Zc3h12a is involved in mRNA decay, but not in TNF regulation, inconsistent with this report.

The Zc3h12a protein has intrinsic RNase activity responsible for the decay of *Il6* mRNA. The mechanism is unique compared to the regulation of other ARE-mediated mRNA decay pathways. For instance, Ttp has been shown to recruit deadenylases for removing polyA tails, facilitating the subsequent degradation of target mRNAs

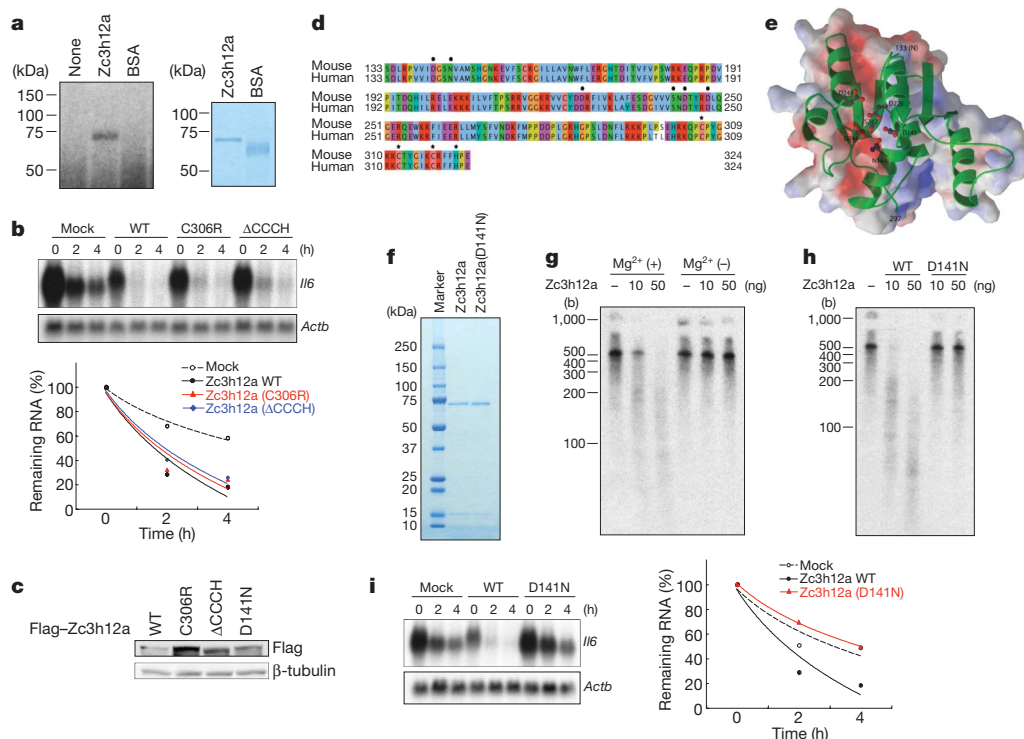


Figure 4 | Zc3h12a contains RNase activity essential for destabilizing *Il6* mRNA. **a**, Binding of Zc3h12a, but not BSA, to *Il6* 3'-UTR (1–403) mRNA by an ultraviolet cross-linking assay. **b**, Role of CCCH zinc-finger motif in destabilizing *Il6* mRNA. HEK293 Tet-off cells were cotransfected with pTREtight-*Il6*-CDS + 3'-UTR, together with expression plasmids encoding Flag-Zc3h12a and its mutants (C306R and Δ CCCH). Then cells were treated with dox for the indicated periods and *Il6* expression was determined by northern blot analysis (top). The autoradiograph was quantified and the ratio of *Il6* to *Actb* was used to determine remaining mRNA levels (bottom). WT, wild type. **c**, Expression levels of Zc3h12a mutant proteins determined by immunoblot. **d**, Alignment of N-terminal and CCCH domains in mouse

and human Zc3h12a. **e**, The structural model of the Zc3h12a N-terminal domain. **f–h**, Zc3h12a possesses RNase activity. The expression levels of synthesized Zc3h12a and Zc3h12a(D141N) are shown (**f**). The RNase activity of Zc3h12a in degrading *Il6* 3'-UTR mRNA (1–403) in the presence or absence of 5 mM Mg^{2+} (**g**). The RNase activity of Zc3h12a and Zc3h12a(D141N) proteins (**h**). Synthesized RNA was incubated with increasing amounts of indicated proteins. RNA size marker is indicated in the left; 'b' denotes bases. **i**, HEK293 Tet-off cells were cotransfected with pTREtight-*Il6* full and Zc3h12a(D141N). The cells were then treated with dox for the indicated periods and *Il6* expression was determined by northern blot analysis.

by exonucleases⁴. Thus, it is intriguing that Zc3h12a has endonuclease activity, which, at least *in vitro*, does not show sequence specificity. The target specificity may be determined by binding partner(s) of Zc3h12a, or Zc3h12a may have a preferential sequence for degradation under certain conditions. The mechanism of how Zc3h12a induces decay of mRNAs is an intriguing topic for further exploration. The RNase domain is conserved in four Zc3h12 family members, and the homologues of this protein family are found in metazoans such as *Drosophila melanogaster* (NCBI accession number CG10889) and *Caenorhabditis elegans* (NCBI accession number C30F12.1). Thus, regulation of mRNA by the RNase domain and CCCH zinc-finger domain seems to be evolutionally conserved.

Another RING-type ubiquitin ligase protein containing a CCCH zinc-finger motif called roquin (also known as Rc3h1) is essential for suppressing autoimmunity by controlling expression of the ICOS costimulatory molecule¹³. Roquin and several microRNAs seem to share an *Icos* 3'-UTR RNA segment for suppressing its degradation¹⁴. Given that each CCCH zinc-finger protein seems to have target mRNA specificity and 60 CCCH-type zinc-finger proteins have been identified in the mammalian genome¹⁵, control of mRNA decay might be as important as the control of transcription in terms of regulation of innate immune responses. Future studies on CCCH zinc-finger proteins will be important for understanding the mechanisms of immune regulation by the control of mRNAs.

METHODS SUMMARY

Mice, reagents, cells and plasmids. Details are given in Methods.

ELISA. IL-4, IL-6, IL-12p40, IL-17, IFN- γ and TNF in culture supernatants, and mouse ANA in serum, were measured by ELISAs following manufacturer's

protocols. ELISAs for mouse IgM, IgG1, IgG2a, IgG2b, IgG3 and anti-double-stranded DNA antibodies in serum were performed as previously described^{16,17}. **Northern blotting, immunoblotting and EMSA.** Northern blotting, immunoblotting and EMSA were performed as previously described¹⁶. **Determination of haematological values.** Haematological analysis of bloods prepared from wild-type and *Zc3h12a*^{-/-} mice were performed at SRL Inc. **Flow cytometry.** Details of flow cytometry are given in Methods. Cells were stained with indicated antibodies. The data were then acquired on a FACS Calibur or FACS Canto II flow cytometer (BD Biosciences), and analysed using FlowJo.

Measurement of RNA stability. mRNA stability was determined using three different methods. First, northern blot analysis was performed using total RNAs prepared from *Zc3h12a*^{-/-} macrophages stimulated with LPS for 2 h followed by treating with actinomycin D. Second, the northern blot Tet-off system was used with HEK293 Tet-off cells transfected by pTRE-*Il6* plasmids together with pFlag-Zc3h12a plasmids. Third, the luciferase assay of cell lysates prepared from HEK293 cells transfected with pGL3 plasmids together with pFlag-Zc3h12a plasmid was used.

In vitro RNA cleavage assay. Cleavage activities of wild-type and mutant forms of Zc3h12a were analysed as described previously¹⁸. After incubation of recombinant Zc3h12a proteins with *in vitro* transcribed [³²P]-labelled RNAs, the cleaved RNAs were purified and analysed by denaturing PAGE and autoradiography. Details of the cleavage assay, expression of recombinant proteins and synthesis of [³²P]-labelled RNAs are given in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 January; accepted 25 February 2009.

Published online 25 March 2009.

1. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* 124, 783–801 (2006).

2. Beutler, B. *et al.* Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. *Annu. Rev. Immunol.* **24**, 353–389 (2006).
3. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **449**, 819–826 (2007).
4. Anderson, P. Post-transcriptional control of cytokine production. *Nature Immunol.* **9**, 353–359 (2008).
5. Barabino, S. M., Hubner, W., Jenny, A., Minvielle-Sebastia, L. & Keller, W. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev.* **11**, 1703–1716 (1997).
6. Kanadia, R. N. *et al.* A muscleblind knockout model for myotonic dystrophy. *Science* **302**, 1978–1980 (2003).
7. Zhao, W., Liu, M. & Kirkwood, K. L. p38 α stabilizes interleukin-6 mRNA via multiple AU-rich elements. *J. Biol. Chem.* **283**, 1778–1785 (2008).
8. Paschoud, S. *et al.* Destabilization of interleukin-6 mRNA requires a putative RNA stem-loop structure, an AU-rich element, and the RNA-binding protein AUF1. *Mol. Cell. Biol.* **26**, 8228–8241 (2006).
9. Datta, S. *et al.* Tristetraprolin regulates CXCL1 (KC) mRNA stability. *J. Immunol.* **180**, 2545–2552 (2008).
10. Taylor, G. A. *et al.* A pathogenetic role for TNF- α in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity* **4**, 445–454 (1996).
11. Zhou, L. *et al.* Monocyte chemoattractant protein-1 induces a novel transcription factor that causes cardiac myocyte apoptosis and ventricular dysfunction. *Circ. Res.* **98**, 1177–1185 (2006).
12. Liang, J. *et al.* A novel CCH-zinc finger protein family regulates proinflammatory activation of macrophages. *J. Biol. Chem.* **283**, 6337–6346 (2008).
13. Vinuesa, C. G. *et al.* A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature* **435**, 452–458 (2005).
14. Yu, D. *et al.* Roquin represses autoimmunity by limiting inducible T-cell co-stimulator messenger RNA. *Nature* **450**, 299–303 (2007).
15. Liang, J., Song, W., Tromp, G., Kolattukudy, P. E. & Fu, M. Genome-wide survey and expression profiling of CCH-zinc finger family reveals a functional module in macrophage activation. *PLoS One* **3**, e2880 (2008).
16. Sato, S. *et al.* Essential function for the kinase TAK1 in innate and adaptive immune responses. *Nature Immunol.* **6**, 1087–1095 (2005).
17. Fukuyama, H., Nimmerjahn, F. & Ravetch, J. V. The inhibitory Fc γ receptor modulates autoimmunity by limiting the accumulation of immunoglobulin G⁺ anti-DNA plasma cells. *Nature Immunol.* **6**, 99–106 (2005).
18. Miyoshi, K., Uejima, H., Nagami-Okada, T., Siomi, H. & Siomi, M. C. *In vitro* RNA cleavage assay for Argonaute-family proteins. *Methods Mol. Biol.* **442**, 29–43 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all colleagues in our laboratory, E. Kamada for secretarial assistance, and Y. Fujiwara, M. Kumagai and R. Abe for technical assistance. We thank S. Sato for discussions and W. Zhao and K. Kirkwood for plasmids. This work was supported by the Special Coordination Funds of the Japanese Ministry of Education, Culture, Sports, Science and Technology, grants from the Ministry of Health, Labour and Welfare in Japan, the Global Center of Excellence Program of Japan, and the NIH (P01 AI070167).

Author Contributions K.M. generated Zc3h12a^{-/-} mice and performed most experiments. O.T. identified Zc3h12a, designed the research and wrote the paper. D.M.S. and H.N. carried out structural modelling. Y.K. analysed microarray data, and T.T. was responsible for histological analysis. T.K., T.M., T.S. and H.K. helped with experiments. S.A. designed the research and supervised the project.

Author Information Microarray data are deposited in the Gene Expression Omnibus (accession number GSE14890 for series of Myd88^{-/-} and Trif^{-/-} macrophages, and GSE14891 for series of Zc3h12a^{-/-} macrophages). The structure model of Zc3h12a nuclease domain has been deposited in the Protein Model DataBase (PMDb) under accession number PM0075640. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.A. (sakira@biken.osaka-u.ac.jp).

METHODS

Generation of *Zc3h12a*^{-/-} mice. Genomic DNA containing *Zc3h12a* was isolated from GSI-I embryonic stem cells and characterized by restriction enzyme mapping and sequencing analysis. A targeting vector was designed to replace exon 3 to exon 5 containing the CCCH type zinc-finger domain, with a neomycin-resistance gene. A 1.1-kb ClaI–BamI fragment was used as the 3' homology, and a 5.9-kb NotI–SalI fragment was used as the 5' homology region. A total of 30 µg of NotI-linearized vector was electroporated into GSI-I embryonic stem cells. After selection with G418, drug-resistant clones were picked up and screened by PCR and Southern blot analysis. These clones were individually microinjected into blastocysts derived from C57BL/6 mice and transferred to pseudopregnant females. Matings of chimaeric male mice to C57BL/6 female mice resulted in the transmission of the mutant allele to the germ line. Resulting *Zc3h12a*^{+/-} mice were intercrossed to generate *Zc3h12a*^{-/-} mice. All animal experiments were done with the approval of the Animal Research Committee of the Research Institute for Microbial Diseases (Osaka University).

Reagents and cells. ELISA kits for mouse IL-4, IL-6, IL-12p40, IL-17, IFN-γ and TNF were purchased from R&D systems. The mouse ANA antibody ELISA kit was purchased from Alpha Diagnostic. Monoclonal anti-YY1 (H-10) and HRP-conjugated monoclonal anti-β-tubulin (D-10) antibodies were obtained from SantaCruz. HRP-conjugated anti-Flag antibody was purchased from Sigma. TLR ligands, including MALP-2, poly(I:C), LPS from *Salmonella Minnesota* Re595, R-848 and CpG oligonucleotide (ODN1668) were obtained as described previously¹⁹.

Peritoneal exudate cells were isolated from the peritoneal cavities of mice 3 days after injection with 2 ml of 4.0% thioglycollate medium (Sigma) by washing with ice-cold Hank's buffered salt solution (Invitrogen). The HEK293 Tet-off cell line was purchased from Clontech.

Plasmids. *Zc3h12a* cDNA was inserted into a pFlag-CMV2 vector (Invitrogen). Point mutations (C306R and D141N) and deletion of the CCCH domain were carried out using the above-mentioned plasmid using QuickChangeII Site-Directed Mutagenesis Kit (Stratagene). pGL3 vector containing full-length (1–403) or parts (1–70, 58–173, 172–403) of *Il6* 3'-UTR sequences were supplied by W. Zhao and K. Kirkwood⁷. Parts (1–92, 1–102, 1–112, 1–132, 1–142 and 122–197) of *Il6* 3'-UTR cDNA were inserted in the pGL3 vector. 3'-UTR cDNA of β-globin (1–130) with or without *Il6* 3'-UTR (77–108) sequence, and the 3'-UTR cDNAs of *Il12p40* (1–781), *Calcr* (1–1601) and *Ifng* (1–631) were inserted in the pGL3 vector. *Il6* CDS and *Il6* CDS + 3'-UTR were inserted in pTREtight vector (Clontech). Wild-type and mutant (D141N) *Zc3h12a* cDNA were inserted in the pGEX-6P1 vector (GE Healthcare). *Il6* 3'-UTR cDNA was inserted downstream of the T7 promoter of pBluescript.

Flow cytometry. Antibodies for flow cytometry were purchased from BD Biosciences. Cell suspensions of spleen were prepared by sieving and gentle pipetting. For surface staining, cells were maintained in the dark at 4 °C throughout. Cells were washed in ice-cold FACS buffer (2% FCS, 0.02% NaN₃ in PBS), then incubated with each antibody for 15 min and washed twice with FACS buffer. Foxp3⁺ regulatory T cell was stained using Mouse Regulatory T Cell Staining Kit (eBioscience) following the manufacturer's instructions. Intracellular cytokines were stained using Cytofix/Cytoparm Plus Fixation/Permeabilization Kit (BD Biosciences) following the manufacturer's instructions. Data were acquired on a FACS Calibur or FACS Canto II flow cytometer (BD Biosciences), and analysed using FlowJo.

Stability of mRNA in macrophages. Peritoneal macrophages (1 × 10⁶) from wild-type and *Zc3h12a*^{-/-} mice were stimulated with LPS (100 ng ml⁻¹) for 2 h. Actinomycin D (2 µg ml⁻¹) was then added to the culture medium to stop transcription, and total RNAs were prepared at the indicated time periods. The RNAs were subjected to northern blot analysis to determine *Il6*, *Tnf*, *Cxcl1* and *Actb* mRNA levels.

The Tet-off system. HEK293 Tet-off cells (3 × 10⁶) were transfected with pTREtight-*Il6*-CDS (which have an *Il6* coding sequence) or pTREtight-*Il6*-CDS + 3'-UTR (which have a *Il6* coding and non-coding 3'-UTR sequence), together with wild-type or mutant forms of *Zc3h12a* expression plasmids or control (empty) plasmid. After 3 h, the cells were subdivided into three 60-mm dishes and cultured overnight. mRNA transcription from pTREtight vectors were terminated by the addition of dox (1 µg ml⁻¹), and total RNA was prepared at the indicated time periods. The RNA was subjected to northern blot analysis to determine *Il6* and *Actb* mRNA levels.

Luciferase assay. HEK293 cells were transfected with pGL3-*Il6* 3'-UTR plasmids or pGL3-empty plasmid together with *Zc3h12a* expression plasmid or empty control plasmid. After 48 h of cultivation, cells were lysed and luciferase activities in the lysates were determined using the Dual-luciferase reporter assay system (Promega). The *Renilla* luciferase gene was simultaneously transfected as an internal control.

Bone marrow transfer. Bone marrow cells were prepared from wild-type and *Zc3h12a*^{-/-} mice. The cells were intravenously injected into lethally irradiated CD45.1 C57BL/6 mice. The chimaeric mice were given neomycin and ampicillin in their drinking water for 4 weeks. The mice were analysed at least 8 weeks after reconstitution. More than 90% of splenocytes from chimaeric mice were CD45.2-positive.

Expression of *Zc3h12a* protein in bacteria. The proteins were expressed in *Escherichia coli* BL21-Gold(DE3)pLysS (Stratagene) transformed with pGEX-6P1-*Zc3h12a* or *Zc3h12a*(D141N) mutant. After expression of the proteins, the cells were collected and resuspended in PBS. The cells were lysed by sonication followed by addition of Triton X-100 at a final concentration of 1% and incubation for 30 min at 4 °C with gentle shaking. The debris were then removed by centrifugation and supernatants were incubated with Glutathione Sepharose 4B (GE Healthcare) for 30 min at 4 °C with gentle shaking. The resins were collected and washed five times with PBS and resuspended in PreScission Protease cleavage buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA and 1 mM dithiothreitol (DTT)). PreScission Protease (GE Healthcare) (80 U) was added and incubated for 4 h at 4 °C with gentle shaking. Supernatants were collected and stored at -80 °C as purified recombinant protein solutions.

Synthesis of [³²P]-labelled RNAs. The pBluescript-*Il6* 3'-UTR (1–430) plasmid was used as a template for the synthesis of RNA having an *Il6* 3'-UTR sequence. *In vitro* RNA synthesis and [³²P] labelling was performed by using Riboprobe *in vitro* Transcription system (Promega) following manufacturer's instructions. The 5'-end labelling was performed using non-labelled RNA and Kinase Max 5'-end labelling Kit (Ambion) following manufacturer's instructions. The 3'-end labelling was performed by incubation of non-labelled RNA with T4 RNA Ligase (Takara) and [³²P]pCp (GE Healthcare).

RNA binding assay. [³²P]-labelled RNA (1 × 10⁶ c.p.m.) were mixed with recombinant protein or BSA (Pierce) in a buffer (25 mM HEPES, 50 mM potassium acetate, 5 mM DTT) and incubated for 20 min at room temperature. Heparin was then added at a final concentration of 0.5 µg ml⁻¹ and incubated further 10 min. The samples were cross-linked by irradiation with 254-nm ultraviolet light using FUNA-UV-LINKER FS-800 (Funakoshi) at a distance of 5 cm from the light source for 20 min on ice. The cross-linked samples were treated with RNaseT (100 U) for 20 min at room temperature, followed by treatment with RNaseA (1 µg) for 15 min at 37 °C. After the digestion, the proteins bound with [³²P]-labelled RNA were analysed by SDS-PAGE and autoradiography.

***In vitro* RNA cleavage assay.** Recombinant proteins and *in vitro* transcribed [³²P]-labelled RNAs (5,000 c.p.m.) were mixed in cleavage buffer (25 mM HEPES, 50 mM potassium acetate, 5 mM DTT) with or without 5 mM magnesium acetate² in the presence of Rnasin plus (40 U) (Promega). The cleaved RNA was purified with Trizol (Invitrogen) and analysed by denaturing PAGE using 6% TBE-urea gel (Invitrogen) and autoradiography.

Microarray analysis. Peritoneal macrophages from wild-type, *MyD88*^{-/-} and *Trif*^{-/-} mice were stimulated with 100 ng ml⁻¹ LPS for 0, 1 and 4 h. Total RNA was extracted with an RNeasy kit (Qiagen), double-stranded cDNA was synthesized from 10 µg of total RNA with the SuperScript Choice System (Invitrogen) primed with T7-(dT) 24 primer. These cDNAs were used to prepare biotin-labelled complementary RNA by an *in vitro* transcription reaction performed using T7 RNA polymerase in the presence of biotinylated ribonucleotides, according to the manufacturer's protocol (Enzo Diagnostics). The cRNA product was purified using an RNeasy kit (Qiagen), fragmented, and hybridized to Affymetrix mouse expression array 430A microarray chips, according to the manufacturer's protocol. For determination of LPS-inducible genes in *Zc3h12a*^{-/-} macrophages, peritoneal macrophages were stimulated with 100 ng ml⁻¹ LPS. Total RNA was then extracted with Trizol (Invitrogen Life Technologies) and further purified using an RNeasy kit. Biotin-labelled cDNA was synthesized from 100 ng of the purified RNA using Ovation Biotin RNA Amplification and Labelling System (Nugen) according to the manufacturer's protocol. Hybridization, staining, washing and scanning of Affymetrix mouse Genome 430 2.0 microarray chip was done following the manufacturer's instructions. Robust multichip average (RMA) expression values were calculated using R and Bioconductor affy package. For hierarchical clustering, probes having a more than two- or fivefold increase compared to 0 h after stimulation were selected. The RMA expression values were transformed to fit averages and standard deviations to zero and one, respectively, by each probe. For analysis of LPS-inducible genes in *MyD88*^{-/-} and *Trif*^{-/-} macrophages, distances between probes were calculated using Pearson's correlation coefficient as a distance function. For analysis of LPS inducible genes in *Zc3h12a*^{-/-} macrophages, principle component analysis for RMA values was performed and Euclidean distances between probes were computed using first to fifth principle components. Hierarchical clustering was carried out using these distances with Ward's method. These calculations and generation of heat map representation were carried out using R and Bioconductor.

Immunohistochemistry. The tissues were fixed with 10% formalin neutral buffer solution, embedded in paraffin, and cut into 5- μ m thick sections. Sections were heated in Target Retrieval Solution (Dako) at 98 °C for 40 min to facilitate antigen retrieval. The sections were incubated with peroxidase-conjugated goat IgG fraction to mouse IgA (alpha chain) (MP Biomedicals) diluted 1:50 by antibody diluent (ChemMate, Dako), or peroxidase-conjugated goat affinity purified F(ab')₂ fragment to mouse IgG (whole molecule) (MP Biomedicals) diluted 1:25 by antibody diluent, for 30 min at room temperature. Immunoreacted cells for mouse IgA and IgG were visualized with diaminobenzidine (Dako). The sections were lightly counterstained with haematoxylin.

Structure modelling. A model of the Zc3h12c N-terminal domain was constructed as follows: first, the sequence was submitted to the BioInfoBank Meta Server (<http://bioinfo.pl>), and the top ten models were built using default settings. The best model was then chosen by submitting each to the SeSAW functional annotation server (<http://pdbs6.pdbj.org/SeSAW/>), and selecting the model with

the highest score. The model chosen was built from the structural genomics template 2qip, using the FFAS03 server (<http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl>) and Modeller²⁰. This model, which also had the highest three-dimensional Jury score, contained a cluster of conserved aspartic acids (D141, D226, S242, D244 and D248) that are also conserved in the active sites of Flap endonucleases (for example, Protein Data Bank ID 1UT5)²¹. Electrostatic surfaces were prepared using the eF-surf server (<http://ef-site.hgc.jp/ef-surf/>) and eF-site²².

19. Kawagoe, T. *et al.* Sequential control of Toll-like receptor-dependent responses by IRAK1 and IRAK2. *Nature Immunol.* **9**, 684–691 (2008).
20. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **Chapter 5**, Unit 5.6 (2006).
21. Feng, M. *et al.* Roles of divalent metal ions in flap endonuclease-substrate interactions. *Nature Struct. Mol. Biol.* **11**, 450–456 (2004).
22. Kinoshita, K. & Nakamura, H. eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics* **20**, 1329–1330 (2004).

The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex

Beom Seok Park¹, Dong Hyun Song¹, Ho Min Kim¹, Byong-Seok Choi¹, Hayyoung Lee³ & Jie-Oh Lee^{1,2}

The lipopolysaccharide (LPS) of Gram negative bacteria is a well-known inducer of the innate immune response¹. Toll-like receptor (TLR) 4 and myeloid differentiation factor 2 (MD-2) form a heterodimer that recognizes a common ‘pattern’ in structurally diverse LPS molecules. To understand the ligand specificity and receptor activation mechanism of the TLR4–MD-2–LPS complex we determined its crystal structure. LPS binding induced the formation of an m-shaped receptor multimer composed of two copies of the TLR4–MD-2–LPS complex arranged symmetrically. LPS interacts with a large hydrophobic pocket in MD-2 and directly bridges the two components of the multimer. Five of the six lipid chains of LPS are buried deep inside the pocket and the remaining chain is exposed to the surface of MD-2, forming a hydrophobic interaction with the conserved phenylalanines of TLR4. The F126 loop of MD-2 undergoes localized structural change and supports this core hydrophobic interface by making hydrophilic interactions with TLR4. Comparison with the structures of tetra-acylated antagonists bound to MD-2 indicates that two other lipid chains in LPS displace the phosphorylated glucosamine backbone by ~5 Å towards the solvent area^{2,3}. This structural shift allows phosphate groups of LPS to contribute to receptor multimerization by forming ionic interactions with a cluster of positively charged residues in TLR4 and MD-2. The TLR4–MD-2–LPS structure illustrates the remarkable versatility of the ligand recognition mechanisms employed by the TLR family^{4,5}, which is essential for defence against diverse microbial infection.

Minute amounts of LPS released from invading bacteria are an early sign of infection and prepare the immune system to counteract further infection¹. They can also lead to fatal septic shock syndrome if the inflammatory response is amplified and uncontrolled. LPS is a glycolipid located in the outer membrane of Gram-negative bacteria. It is composed of an amphipathic lipid A component and hydrophilic polysaccharides of the core and O-antigen^{6,7}. Lipid A represents the conserved molecular pattern of LPS and is the main inducer of immunological responses to LPS. TLR4 in association with MD-2 is responsible for the physiological recognition of LPS^{8,9}. So far, ten members of the TLR family, recognizing a wide variety of microbial products, have been identified in humans¹⁰. The extracellular domains of TLRs consist of leucine-rich repeats (LRRs) with a horseshoe-like shape^{2,4,11,12}. Binding of agonistic ligands causes dimerization of the extracellular domains and is believed to trigger the recruitment of specific adaptor proteins to the intracellular domains, thus initiating a signalling cascade^{4,5} (Supplementary Fig. 1). LPS is extracted from the bacterial membrane and transferred to TLR4–MD-2 by two accessory proteins, LPS-binding protein and CD14¹³. The TLR4–MD-2 heterodimer has complex ligand specificity. It can be activated by structurally diverse LPS molecules, and apparently minor changes in synthetic

derivatives of LPS can abolish their endotoxic potency^{7,14,15}. The lack of a high-resolution structure is in part responsible for incomplete understanding of the basis of receptor specificity and of the activation mechanism. We have therefore determined the crystal structure of the TLR4–MD-2–LPS complex at 3.1 Å resolution.

The receptor multimer is composed of two copies of the TLR4–MD-2–LPS complex arranged in a symmetrical fashion (Fig. 1a).

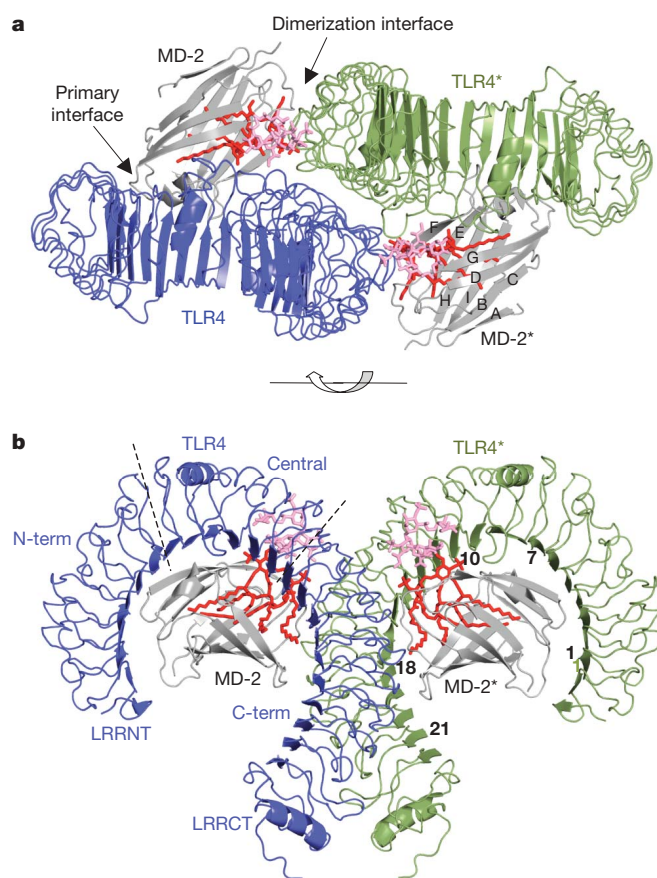


Figure 1 | Overall structure of the TLR4–MD-2–LPS complex. **a**, Top view of the symmetrical dimer of the TLR4–MD-2–LPS complex. The primary interface between TLR4 and MD-2 is formed before binding LPS, and the dimerization interface is induced by binding LPS. **b**, Side view of the complex. The lipid A component of LPS is coloured red, and the inner core carbohydrates of LPS are coloured pink. The module numbers of the LRRs in TLR4 and the names of the β strands in MD-2 are written in black. TLR4 is divided into N-, central and C-terminal domains². The LRRNT and LRRCT modules cover the amino and carboxy termini of the LRR modules.

¹Department of Chemistry and ²Institute for the BioCentury, KAIST, Daejeon, 305-701, Korea. ³Department of Biology, School of Bioscience & Biotechnology, Chungnam National University, Daejeon, 305-764, Korea.

Comparison with the recently published structure of the monomeric TLR4 and MD-2 complex shows that the overall folding of TLR4 and MD-2 is not disturbed by LPS binding or dimerization². TLR4 adopts the characteristic horseshoe-like shape of the LRR superfamily¹⁶ (Fig. 1b). MD-2 has a β -cup fold structure composed of two anti-parallel β sheets forming a large hydrophobic pocket for ligand binding^{2,3}. LPS binds to this pocket and directly mediates dimerization of the two TLR4–MD-2 complexes (Fig. 1a). As shown previously², the primary contact interface between TLR4 and MD-2 that is formed before LPS binding involves two chemically distinct regions, the A and B patches provided by the N-terminal and the central domains of TLR4, respectively. The main dimerization interface of MD-2 is located on the opposite side of the primary interface and interacts with LRR modules 15–17 in the C-terminal domain of TLR4 (Fig. 1a and Supplementary Fig. 2a). For dimerization, TLR4 forms hydrophobic and hydrophilic bonds directly with LPS and the surrounding F126 and L87 loops of MD-2. The F126 and L87 loops connect the β G– β H and β E– β F strands of MD-2, respectively (Supplementary Fig. 2b). Throughout this paper, we have marked the second TLR4 and MD-2 molecules and their amino acid residues in the heterotetrameric complex with an asterisk to distinguish them from those of the primary TLR4 and MD-2 complex (Fig. 1).

The lipid A component of *Escherichia coli* LPS contains two phosphorylated glucosamines that are connected by a β (1–6) linkage and acylated by six lipid chains (Fig. 2a). The lipid chain attached to the C2' carbon of the glucosamine is labelled R2'. The R2'' chain is connected

to the R2' lipid by an ester link. The other lipid chains are named in a similar manner. In the crystal structure these lipid chains interact with the hydrophobic pocket in MD-2 (Fig. 2b). The carbon chains of lipids R3, R2', R3', R2'' and R3'' are completely buried inside the pocket, but the R2 chain is partially exposed to the MD-2 surface composing the core hydrophobic interface for interaction with TLR4*. The ester and amide groups connecting the lipids to the glucosamine backbone or to the other lipid chains are exposed to the surface of MD-2. They interact with hydrophilic side chains located on the β G strand of the MD-2 pocket and on the surface of TLR4 and TLR4* (Fig. 2c). The two phosphate groups of the lipid A bind to the TLR4–MD-2 complex by interacting with positively charged residues in TLR4, TLR4* and MD-2 and making a hydrogen bond to S118 of MD-2 (Fig. 2d).

Both hydrophobic and hydrophilic interactions contribute to the main dimerization interaction between MD-2/LPS and TLR4* (Fig. 3a). The hydrophobic R2 lipid chain of LPS interacts directly with a small hydrophobic patch on the surface of TLR4* composed of two phenylalanines, F440* and F463* at the core, and L444* at the periphery (Fig. 3b, c). The hydrophobic residues V82, M85, L87, I124 and F126 of MD-2 supplement this core hydrophobic interface. Hydrophilic residues in the F126 loop and the R90 residue of MD-2 form hydrogen bonds and ionic interactions with TLR4* that surround and support the hydrophobic core of the dimerization interface. The 3-hydroxyl group of the R2 chain of LPS contributes to the interaction by forming a hydrogen bond with Q436* of TLR4*. Two phosphate groups, the 1-phosphate and 4'-phosphate of lipid A, also

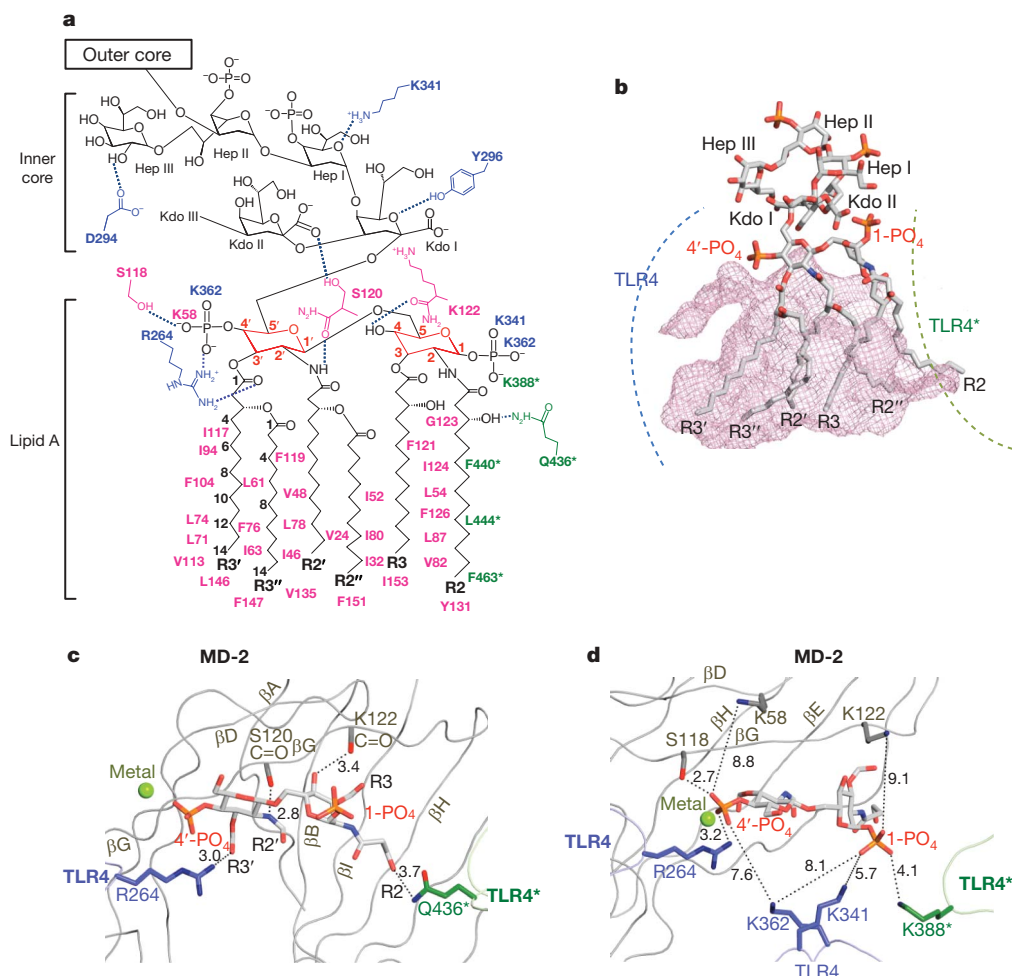


Figure 2 | Binding of LPS to TLR4 and MD-2. **a**, Chemical structure of the Ra chemotype of *E. coli* LPS²⁴. The lipid chains are labelled. The carbons of the glucosamines and lipid chains of lipid A are numbered. Hydrogen bonds are shown by broken blue lines. **b**, The molecular surface of the MD-2 pocket is

drawn in mesh. **c**, Hydrogen bonds between lipid A and TLR4–MD-2. **d**, Ionic and hydrogen bond interactions of the two phosphate groups of lipid A. Interaction distances in Ångströms are written. Inner core carbohydrates and carbon chains of the lipids are omitted for clarity.

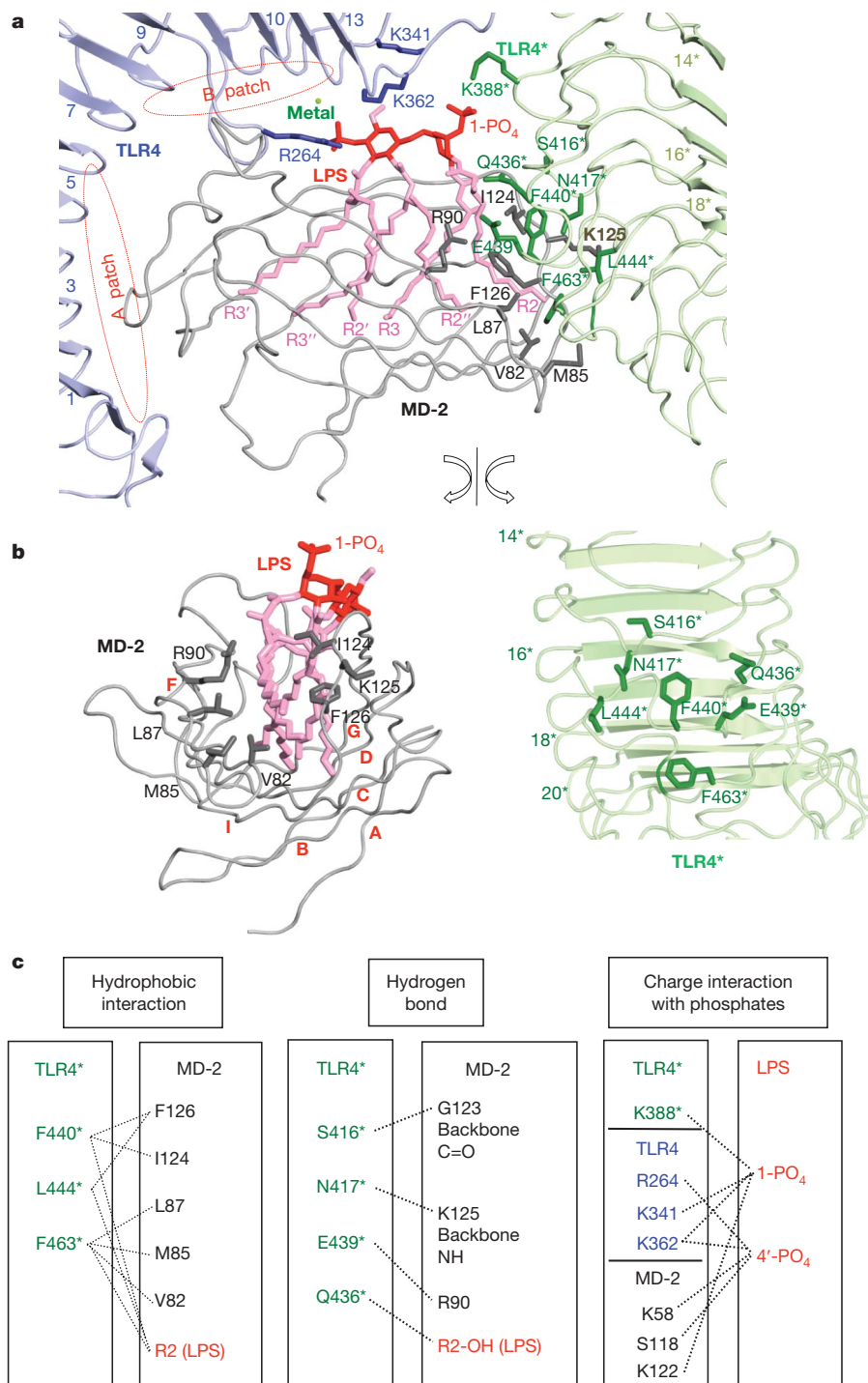


Figure 3 | The main dimerization interface of the TLR4-MD-2-LPS complex. **a**, Overall shape of the main dimerization interface. The inner core of LPS is omitted for clarity. The primary interface is classified into A and B patches, which are marked in red. K58, S118 and K122 of MD-2 are not shown for clarity. **b**, The dimerization interface has been split and rotated to

show the residues involved. The LRR module numbers of TLR4* and the β strands of MD-2 are labelled in green and red, respectively. **c**, Residues involved in the core dimerization interface. The interaction partners are connected by broken lines.

play an important role in dimerization by binding to a positively charged cluster of lysines and an arginine from TLR4, TLR4* and MD-2 (Figs 2d and 3a, c). LPS binding and dimerization induce localized changes in the structure of the F126 loop of MD-2, and in the radius and bending angle of TLR4 (Supplementary Figs 3 and 4).

In addition to the main interaction described above, several secondary interactions contribute to the dimerization. A metal ion found near the 4'-phosphate group appears to connect the MD-2-LPS complex to TLR4 either directly, or indirectly through water

molecules (Supplementary Fig. 5). This is probably a magnesium ion supplied by the crystallization buffer and appears to be dispensable for LPS binding and receptor dimerization because neither are blocked by EDTA treatment (data not shown). TLR4 makes an additional contribution to dimerization by directly contacting TLR4* (Supplementary Fig. 6). Carbohydrates in the inner core of LPS form several bonds with the metal ion and the MD-2 and TLR4 molecules (Supplementary Fig. 7). The importance of these carbohydrate interactions in receptor dimerization is, however, not obvious, because

they are not essential for the endotoxic activity of LPS¹⁷. The Kdo III group of the inner core and all carbohydrates in the outer core are not clearly visible in the electron density map, presumably because their structures and positions are not fixed by interaction with the proteins.

Our structural observations on the TLR4–MD-2–LPS complex are supported by previous biochemical and mutagenesis studies. We have shown that truncation of the C-terminal domain of TLR4 blocks LPS-induced dimerization without compromising LPS binding activity². In the crystal structure, the main dimerization interface of TLR4 is located at LRR modules 15–17 in the C-terminal domain. Therefore, truncation of all these LRR modules should completely block dimerization. Mutation of F126 and the surrounding residues in MD-2 also blocked LPS-induced dimerization^{2,18}. Residue F126 is located in the core of the dimerization interface and initiates structural changes in MD-2 (Fig. 3 and Supplementary Fig. 3). In the monomeric structures of MD-2 bound to non-agonistic ligands^{2,3}, F126 is exposed to the solvent area without interacting with any ligand or protein residues (Supplementary Fig. 3b). On the contrary, in the dimerized structure of the TLR4–MD-2–LPS complex, F126 together with L54, Y131 and I124 of MD-2 forms extensive hydrophobic bonds with lipid chains R2 and R3 of LPS, and F440* of TLR4* (Supplementary Fig. 3c). These interactions are important for positioning the R2 lipid chain correctly and inducing a ~5 Å structural shift in the F126 loop, which moves the critical residues G123, I124 and K125 to positions suitable for the dimerization interaction with TLR4*. Hence, changing F126 to alanine, with its smaller side chain, should disrupt this core interaction and prohibit receptor dimerization. Mutations of other residues in the F126 loop also interfere with LPS binding and signalling^{19–22}. Interestingly, a mutant MD-2 with K125 changed to alanine shows normal LPS binding and receptor activation activity²². This is because the backbone atoms, but not the side chain atoms, are involved in receptor dimerization.

Recently the structures of MD-2 bound by two antagonists, Eritoran and lipid IVa, have been reported^{2,3}. In these crystal structures, the four lipid chains of the antagonists completely fill the available space in the pocket. *E. coli* LPS has two more lipid chains

than these antagonists, so it has been proposed that global structural changes may take place in the MD-2 pocket to accommodate the extra lipid chains^{2,3,23}. Unexpectedly, the crystal structure of the TLR4–MD-2–LPS complex demonstrates that the size of the MD-2 pocket is unchanged and that additional space for lipid binding is generated by displacing the glucosamine backbone upwards by ~5 Å (Fig. 4). This shift of the glucosamines repositions the phosphate groups such that they can interact with positively charged residues of TLR4 and TLR4*, thus promoting the dimerization and activation of the receptor complex (Figs 2d and 3a, c). Interestingly, the glucosamine backbones of the antagonists are not only translated but also rotated by ~180 degrees, so interchanging the two phosphate groups. More research is needed to establish whether all LPS derivatives with four lipid chains cause a similar rotation of the phospho-glucosamine backbones.

The structure-activity relationship of the lipid A of LPS has been extensively studied and several factors governing the immunological activity of LPS have been identified^{15,24}. Of these, the total number of lipid chains is the most important factor. Lipid A with six lipid chains has optimal inflammatory activity, while lipid As with five lipid chains are ~100-fold less active, and those with four lipid chains, such as Eritoran, completely lack agonistic activity^{25,26}. The lipid chains of LPS interact hydrophobically with MD-2, and hydrophobic interactions are not sensitive to distance and angle, so if the chemical structure of the chains is changed their positions can be shifted to maximize hydrophobic contact. Therefore, in LPS with five or less lipid chains, all the lipid chains probably move further into the pocket to fill the empty space, and there should be substantial energetic penalties when they move back to the surface of MD-2 for dimerization with TLR4*.

The two phosphate groups in the lipid A region also greatly affect the endotoxic activity of LPS^{15,24}. Deletion of either of these phosphate groups reduces endotoxic activity ~100-fold and the resulting monophosphoryl LPS (MPL) is only a weak activator of the human innate immune response. Furthermore, an MPL based on lipid A from *Salmonella minnesota* selectively activates the TLR4–TRAM–TRIF signalling pathway but not the TLR4–Mal–MyD88 pathway²⁷ (Supplementary Fig. 1). This data suggest that deletion of the 1-phosphate

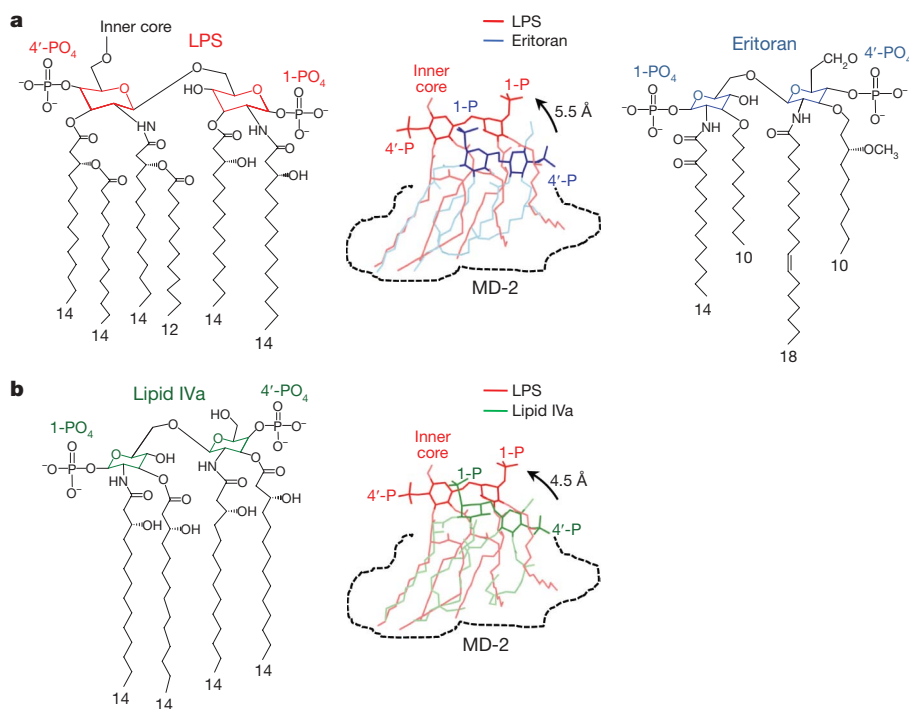


Figure 4 | Structural comparison of LPS with antagonists. **a**, Comparison of LPS and Eritoran after superimposition of MD-2. LPS and Eritoran are coloured red and blue, respectively. **b**, Comparison of LPS and lipid IVa. The structures of LPS and lipid IVa are shown after superimposition of MD-2.

The shape of the MD-2 pocket is drawn schematically with broken lines. The glucosamine rings and the phosphate groups are represented in darker colours.

not only weakens the ligand affinity but also induces structural rearrangement of the TLR4–MD-2–adaptor multimer. In the crystal structure, the 1- and 4'-phosphate groups interact with a cluster of positively charged residues from TLR4, TLR4* and MD-2 (Fig. 2d and 3a, c). Therefore their deletion should have a profound impact on the receptor–ligand interaction. In natural LPSs, chemical groups such as 2-aminoethylphosphate or L-Ara4N are occasionally attached to the phosphate groups of lipid A²⁴. Previous studies have also shown that removal of the negative charges of the phosphate groups by chemical modification dramatically reduces the potency of LPS, whereas replacing the phosphates with other negatively charged groups such as phosphonoxyethyls has only minor effects^{15,24,28}. This is because the 1- and 4'-phosphates in the TLR4–MD-2–LPS complex form mostly medium-range ionic bonds, and there is space for extra atoms in the binding sites (Fig. 2d). Therefore, substituting other negatively charged groups for the phosphates should not always disrupt binding, although it may modulate the activity of the resulting complexes. Domain swapping data involving horse and human TLR4–MD-2s further support the importance of the phosphate groups in LPS recognition²⁹ (Supplementary Discussion). The structure of the TLR4–MD-2–LPS complex illustrates the remarkable variety of the ligand recognition mechanisms of the TLR family^{4,5} (Supplementary Discussion).

In summary, the crystal structure shows that multiple structural components of the TLR4–MD-2 receptor are involved in LPS recognition. Further work is required to establish how these components respond to the structurally diverse LPS variants from different bacterial species. Such studies may lead to the discovery of novel LPS derivatives with desirable therapeutic properties.

METHODS SUMMARY

The ectodomain of human TLR4 (amino acids 27–631) and the human MD-2 gene (amino acids 19–160) fused to the Protein A tag were co-expressed in High Five insect cells (Invitrogen) and purified by IgG Sepharose affinity chromatography (GE Healthcare Life Sciences). A list of the polymerase chain reaction (PCR) primers used for cloning is given in Supplementary Table 1. The Protein A tag was removed with thrombin. The TLR4–MD-2 complex was further purified by SP Sepharose ion exchange and Superdex 200 gel filtration chromatography (GE Healthcare Life Sciences). The purified TLR4–MD-2 complex was concentrated to ~1 mg ml⁻¹ and treated with PNGase F. The Ra chemotype of *E. coli* LPS (Sigma, L9641) was incubated with the partially deglycosylated TLR4–MD-2 complex. After the reaction, the TLR4–MD-2–LPS complex was purified by Superdex 200 gel filtration chromatography to remove unbound LPS Ra. Crystals of the complex appeared after three days at 5 °C, when we employed the hanging-drop vapour diffusion method, mixing 1 µl of protein solution and 1 µl of crystallization buffer containing 50 mM MgCl₂, 0.1 M Na-HEPES pH 7.5 and 30% PEG MME 550. Initial phases were calculated by molecular replacement using the program PHASER³⁰ (Supplementary Table 2). The atomic models were refined by iterative modelling and refinement cycles. LPS was built into a strong and continuous electron density found in the MD-2 pocket (Supplementary Fig. 8).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 November 2008; accepted 27 January 2009.

Published online 1 March 2009.

- Beutler, B. & Rietschel, E. T. Innate immune sensing and its roots: the story of endotoxin. *Nature Rev. Immunol.* **3**, 169–176 (2003).
- Kim, H. M. *et al.* Crystal structure of the TLR4–MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* **130**, 906–917 (2007).
- Ohto, U., Fukase, K., Miyake, K. & Satow, Y. Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science* **316**, 1632–1634 (2007).
- Jin, M. S. *et al.* Crystal structure of the TLR1–TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130**, 1071–1082 (2007).
- Liu, L. *et al.* Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* **320**, 379–381 (2008).
- Raetz, C. R. Biochemistry of endotoxins. *Annu. Rev. Biochem.* **59**, 129–170 (1990).
- Raetz, C. R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700 (2002).
- Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
- Shimazu, R. *et al.* MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J. Exp. Med.* **189**, 1777–1782 (1999).
- Matsushima, N. *et al.* Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genom.* **8**, 124 (2007).
- Bell, J. K. *et al.* The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proc. Natl Acad. Sci. USA* **102**, 10976–10980 (2005).
- Choe, J., Kelker, M. S. & Wilson, I. A. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* **309**, 581–585 (2005).
- Miyake, K. Roles for accessory molecules in microbial recognition by Toll-like receptors. *J. Endotoxin Res.* **12**, 195–204 (2006).
- Erridge, C., Bennett-Guerrero, E. & Poxton, I. R. Structure and function of lipopolysaccharides. *Microbes Infect.* **4**, 837–851 (2002).
- Rietschel, E. T. *et al.* Bacterial endotoxin: molecular relationships of structure to activity and function. *FASEB J.* **8**, 217–225 (1994).
- Bella, J., Hindle, K. L., McEwan, P. A. & Lovell, S. C. The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **65**, 2307–2333 (2008).
- Galanos, C. *et al.* Synthetic and natural *Escherichia coli* free lipid A express identical endotoxic activities. *Eur. J. Biochem.* **148**, 1–5 (1985).
- Kobayashi, M. *et al.* Regulatory roles for MD-2 and TLR4 in ligand-induced receptor clustering. *J. Immunol.* **176**, 6211–6218 (2006).
- Kawasaki, K., Nogawa, H. & Nishijima, M. Identification of mouse MD-2 residues important for forming the cell surface TLR4–MD-2 complex recognized by anti-TLR4–MD-2 antibodies, and for conferring LPS and taxol responsiveness on mouse TLR4 by alanine-scanning mutagenesis. *J. Immunol.* **170**, 413–420 (2003).
- Re, F. & Strominger, J. L. Separate functional domains of human MD-2 mediate Toll-like receptor 4-binding and lipopolysaccharide responsiveness. *J. Immunol.* **171**, 5272–5276 (2003).
- Visintin, A., Latz, E., Monks, B. G., Espevik, T. & Golenbock, D. T. Lysines 128 and 132 enable lipopolysaccharide binding to MD-2, leading to Toll-like receptor-4 aggregation and signal transduction. *J. Biol. Chem.* **278**, 48313–48320 (2003).
- Teghanemt, A. *et al.* Novel roles in human MD-2 of phenylalanines 121 and 126 and tyrosine 131 in activation of Toll-like receptor 4 by endotoxin. *J. Biol. Chem.* **283**, 1257–1266 (2008).
- Jin, M. S. & Lee, J. O. Structures of the Toll-like receptor family and its ligand complexes. *Immunity* **29**, 182–191 (2008).
- Rietschel, E. T. *et al.* The chemical structure of bacterial endotoxin in relation to bioactivity. *Immunobiology* **187**, 169–190 (1993).
- Teghanemt, A., Zhang, D., Levis, E. N., Weiss, J. P. & Gioannini, T. L. Molecular basis of reduced potency of underacylated endotoxins. *J. Immunol.* **175**, 4669–4676 (2005).
- Rosignol, D. P. & Lynn, M. TLR4 antagonists for endotoxemia and beyond. *Curr. Opin. Investig. Drugs* **6**, 496–502 (2005).
- Mata-Haro, V. *et al.* The vaccine adjuvant monophosphoryl lipid A as a TRIF-biased agonist of TLR4. *Science* **316**, 1628–1632 (2007).
- Ulmer, A. J. *et al.* Biological activity of synthetic phosphonoxyethyl analogs of lipid A and lipid A partial structures. *Infect. Immun.* **60**, 3309–3314 (1992).
- Walsh, C. *et al.* Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa. *J. Immunol.* **181**, 1245–1254 (2008).
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of beamline 4A at the Pohang Accelerator Laboratory and beamline ID23-2 at ESRF for help with data collection. We thank J. Gross for critical reading of the manuscript. J.-O.L. and co-workers are funded by the Creative Research Initiative (Center for Membrane Receptor Research) from the Ministry of Education, Science and Technology of Korea.

Author Contributions B.S.P., D.H.S. and H.M.K. performed the experiments. B.S.P. and J.-O.L. designed the experiments and analysed the data. B.S.P., H.L. and J.-O.L. wrote the paper. J.-O.L. managed the project and had overall responsibility for data interpretation and writing the manuscript. All authors discussed and commented on the manuscript.

Author Information Atomic coordinates and the structure factor files have been deposited in the Protein Data Bank (<http://www.rcsb.org>) under accession number 3FXI. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.-O.L. (ijeoh.lee@kaist.ac.kr).

METHODS

Protein expression and purification. The ectodomain of human TLR4 (amino acids 27–631) was cloned into the BamHI and NotI sites of pAcGP67 vector (BD Biosciences). The human MD-2 gene (amino acids 19–160) was fused to the Protein A gene derived from pEZZ18 (GE Healthcare Life Sciences) and cloned into the BamHI and NotI sites of pAcGP67 vector (BD Biosciences). A thrombin cleavage site was introduced between the MD-2 and the Protein A genes for removal of the tag. A list of the PCR primers used for cloning is given in Supplementary Table 1. The amino acid sequences of the cloned genes were confirmed by DNA sequencing. TLR4 and MD-2 were co-expressed in High Five insect cells (Invitrogen) and purified by IgG Sepharose affinity chromatography (GE Healthcare Life Sciences). The Protein A tag was removed with thrombin. For the digestion, the concentration of the TLR4–MD-2 complex was set at 1–2 mg ml^{−1}, and 0.1% (w/w) thrombin was used at 4 °C overnight. The TLR4–MD-2 complex was further purified by SP Sepharose ion exchange and Superdex 200 gel filtration chromatography (GE Healthcare Life Sciences). The purified TLR4–MD-2 complex was concentrated to ~1 mg ml^{−1} and treated with PNGase F at 37 °C for 3 h in a buffer containing 20 mM Tris pH 8.0 and 200 mM NaCl. The molar ratio of the protein to PNGase F was 3:1. The partially deglycosylated TLR4–MD-2 complex was purified by SP Sepharose ion exchange and Superdex 200 gel filtration chromatography.

LPS binding and crystallization. To screen LPS chemotypes suitable for crystallization experiments, we tested the binding and dimerizing activities of the Ra, Rc and Re forms of LPS and lipid A. A buffer solution containing Ra or Rc, the LPS chemotypes with longer carbohydrate chains, became transparent after sonication, suggesting that these LPS molecules form small and homogeneous micelles. After incubation with these LPS types in our reaction condition, ~80% of TLR4–MD-2 protein bound to the LPS and was dimerized. On the other hand, a buffer solution containing lipid A or LPS Re with shorter carbohydrate chains, remained turbid even after sonication, suggesting that these LPS molecules form larger and highly heterogeneous aggregates owing to their greater hydrophobicity versus hydrophilicity ratio. As a result, binding of these LPS types by TLR4–MD-2 is significantly less efficient, probably because of the lower efficiency with which they are extracted from the large aggregates. We did not examine this problem systematically, but simple addition of mouse CD14 and LPS-binding protein did not make any significant difference to the binding efficiency of LPS under our reaction conditions. For crystallization, the Ra chemotype of *E. coli* LPS (Sigma, L9641) was sonicated for 10 min and incubated with the deglycosylated TLR4–MD-2 complexes at 37 °C for 3 h. The reaction buffer contained 20 mM Tris-HCl, pH 8.0 and 200 mM NaCl. The concentration of TLR4–MD-2

was ~1 mg ml^{−1} and the molar ratio of LPS to the protein was 10:1. In the Ra chemotype of LPS, the O antigen carbohydrates of LPS are absent but the lipid A and inner and outer cores remain intact⁶. After the reaction, the TLR4–MD-2–LPS complex was purified by Superdex 200 gel filtration chromatography to remove unbound LPS Ra. Binding of LPS Ra to the protein complex was monitored by 8% native polyacrylamide gel electrophoresis (PAGE) and gel filtration chromatography. The TLR4–MD-2–LPS complex was concentrated to 2 mg ml^{−1} in a buffer containing 20 mM Tris pH 8.0 and 200 mM NaCl. Crystals of the complex appeared after three days at 5 °C, when we employed the hanging-drop vapour diffusion method mixing 1 µl of protein solution and 1 µl of crystallization buffer containing 50 mM MgCl₂, 0.1 M Na-HEPES pH 7.5 and 30% PEG MME 550.

Data collection. For data collection, crystals were flash-frozen at −170 °C in the crystallization buffer with a 3% higher concentration of PEG MME 550. Diffraction data were collected at the 4A beamline of the Pohang Accelerator Laboratory and the ID23-2 beamline of the European Synchrotron Radiation Facility. The XDS package was used to index, integrate and scale the diffraction data (Supplementary Table 2)³¹.

Model building and refinement. Initial phases were calculated by molecular replacement using the program PHASER³⁰. The published structures of TV3–MD-2–Eritoran and the VT3 hybrid were used together as search probes². Model building was aided by the structures of mouse TLR4–MD-2 and the TV8 hybrid. The atomic models were refined by iterative modelling and refinement cycles using programs O and CNS^{32,33}. The atomic coordinates of the model were restrained by two-fold non-crystallographic symmetry throughout the refinement. The atomic model of LPS was built into a strong and continuous electron density found in the MD-2 pocket (Supplementary Fig. 8). The final structure was refined to *R* and *R*_{free} factors of 24% and 28%, respectively (Supplementary Table 2). 87% of protein residues are present in the favoured regions and no non-glycine residues are in the disallowed regions of the Ramachandran plots³⁴.

31. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
32. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
33. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
34. Lovell, S. C. *et al.* Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* **50**, 437–450 (2003).

LETTERS

A mutation in *Ihh* that causes digit abnormalities alters its signalling capacity and range

Bo Gao^{1,2*}, Jianxin Hu^{1,2*}, Sigmar Stricker^{3,4}, Martin Cheung^{1,5}, Gang Ma², Kit Fong Law¹, Florian Witte^{3,4,6}, James Briscoe⁷, Stefan Mundlos^{3,4}, Lin He^{2,8,9}, Kathryn S. E. Cheah^{1,5} & Danny Chan^{1,5}

Brachydactyly type A1 (BDA1) was the first recorded disorder of the autosomal dominant Mendelian trait in humans, characterized by shortened or absent middle phalanges in digits. It is associated with heterozygous missense mutations in indian hedgehog (*IHH*)^{1,2}. Hedgehog proteins are important morphogens for a wide range of developmental processes^{3,4}. The capacity and range of signalling is thought to be regulated by its interaction with the receptor PTCH1 and antagonist HIP1. Here we show that a BDA1 mutation (E95K) in *Ihh* impairs the interaction of IHH with PTCH1 and HIP1. This is consistent with a recent paper showing that BDA1 mutations cluster in a calcium-binding site essential for the interaction with its receptor and cell-surface partners⁵. Furthermore, we show that in a mouse model that recapitulates the E95K mutation, there is a change in the potency and range of signalling. The mice have digit abnormalities consistent with the human disorder.

Mutations in human *SHH* and *IHH* cause developmental defects such as holoprosencephaly⁶, acrocapitofemoral dysplasia⁷ and BDA1 (refs 1, 2). However, the molecular basis of these mutations is not known. To investigate the mechanism underlying BDA1 *in vivo*, we generated mice with the *Ihh*^{E95K} mutation by gene targeting (Supplementary Fig. 1). *Ihh*^{+/E95K} mice showed shortened middle phalanges in digits II and V, indicating a dominant effect of the mutation on digit formation, because mice heterozygous for an *Ihh* null mutation have normal digits (Fig. 1a and Supplementary Fig. 2a). Homozygous mice (*Ihh*^{E95K/E95K}) showed a classic BDA1 phenotype with severely shortened (II–IV) or absent middle phalanges (digit V), and abnormal and shortened limb skeletal elements from embryonic day (E)13.5 (Supplementary Fig. 2b–d). Because heterozygous mutants are only mildly affected, subsequent analyses were performed on homozygous mutants, referred to as BDA1 mice.

In limb development, *SHH* acts early, regulating patterning and growth⁸. *IHH* acts later and is thought not to affect patterning, but to regulate endochondral bone formation by controlling chondrogenic differentiation and proliferation⁹. The mammalian growth plate is an ideal system to study IHH signalling because it is organized into distinct zones that progress from round chondrocytes to flattened proliferating cells organized into columns, to larger non-dividing prehypertrophic cells that express *Ihh*, and finally to even larger hypertrophic chondrocytes expressing *Col10a1* (ref. 9). Endochondral ossification in the mutants was retarded, as indicated by the delayed expression of *Ihh* and *Col10a1* in the developing ulna and radius of heterozygous and homozygous mice at E14.5 (Supplementary Fig. 2e).

The range and capacity of IHH signalling can be monitored by the expression of hedgehog targets *Ptch1* and *Gli1*. In wild-type mice at E15.5, the expression of *Ptch1* (Fig. 1b and Supplementary Fig. 3) and *Gli1* (data not shown) decreased gradually from the source of IHH (prehypertrophic chondrocytes) to the proliferative zone. In BDA1 mice, *Ptch1* expression in the proliferating zone was initially reduced and the gradient faded more quickly across the proliferative zone than in wild-type mice, indicating a change in the IHH activity gradient. IHH promotes chondrocyte proliferation in the growth plate¹⁰. Pulse-labelling with BrdU at E15.5 revealed fewer BrdU-positive chondrocytes in the developing cartilage anlagen of the limb bones in BDA1 mice (Fig. 1c, d), consistent with a reduced capacity of IHH to stimulate chondrocyte proliferation. Quantification of the IHH–PTCH1 binding affinity using a cell-based assay¹¹ showed a higher dissociation constant (K_d) for IHH(E95K) (40.6 nM) than IHH (20.6 nM), indicating a markedly reduced binding affinity of IHH(E95K) (Fig. 1e) that could explain the reduced signalling capacity. Furthermore, analysis from an *in vitro* cell assay showed that the signalling capacity of IHH(E95K) was reduced to about 70% of wild type (Supplementary Fig. 4).

In the growth plate, there is also a ‘long-range’ IHH signal to the periarticular region where it activates the expression of *PthrP* (also known as *Pthlh*) as part of the IHH–PTHrP negative feedback regulatory loop^{10,12}. Notably, in BDA1 mice, despite the reduced ‘short-range’ hedgehog signalling, periarticular expression of *Ptch1* (Fig. 1b) and *PthrP* (Supplementary Fig. 3a) was increased, indicating that IHH(E95K) had travelled further than normal. CDO (also known as CDON) is a positive regulator of hedgehog signalling¹³, whereas PTCH1 and HIP1 are negative regulators¹⁴. Hedgehog induces expression of *Ptch1* and *Hip1* to form a feedback loop that restricts hedgehog signalling⁴. A plausible explanation for the altered signalling in BDA1 mice is that an impaired interaction with PTCH1 and other signal transducers decreases the signalling efficacy of IHH but also decreases its retention close to its source allowing it to diffuse further. An alternative is that the mutation affects the multimerization and stability of IHH. However, this is not known and we favour the former because the signalling capacity of IHH(E95K) is reduced in C3H10T1/2 cells (Supplementary Fig. 4), and BDA1 missense mutations in IHH affect its interactions with PTCH1, HIP and CDO⁵.

We studied further the effect of the E95K mutation on digit development. In the developing digit, *Hip1* is expressed at the periphery of the condensed cartilage elements, as well as in the proximal and distal margins of adjacent cartilage elements flanking the interzone—the site

¹Department of Biochemistry, the University of Hong Kong, Hong Kong, China. ²Bio-X Center, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200030, China. ³Max-Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany. ⁴Institut für Medizinische Genetik, Charité, Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. ⁵Centre for Reproduction, Development and Growth, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, China. ⁶Institute for Chemistry/Biochemistry, Free University Berlin, Thielallee 63, 14195 Berlin, Germany. ⁷Developmental Neurobiology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. ⁸Institute for Nutritional Sciences, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. ⁹Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China.

*These authors contributed equally to this work.

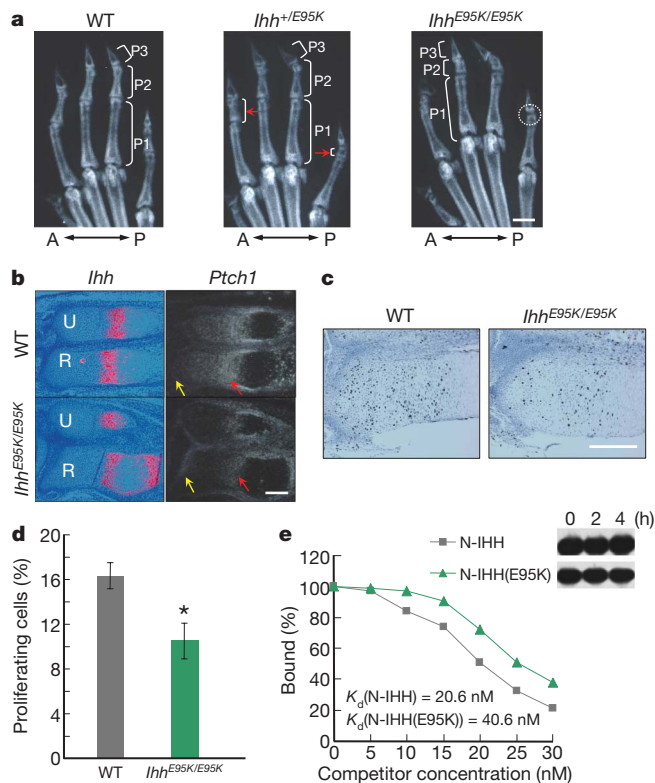


Figure 1 | E95K mutation in IHH altered the capacity and range of hedgehog signalling in BDA1 mouse. **a**, Digit radiography of 7-weeks-postnatal mice. Digits I–V are numbered from anterior (A) to posterior (P). The middle phalanges of digits II and V were shortened in *Ihh*^{E95K/E95K} mice (red arrow). In *Ihh*^{+/E95K} mice, the middle phalanges were all shortened in digits II–IV and missing in digit V (dotted circle). P1, proximal phalange; P2, middle phalange; P3, distal phalange; WT, wild type. **b**, In the wild type, *Ptch1* expression relative to the *Ihh* expression domain showed a gradient with a gradual decrease from the IHH source. In contrast, *Ptch1* expression decreased markedly from the IHH source in the *Ihh*^{E95K/E95K} mouse growth plate (red arrow regions), but was maintained/increased in the periarticular region (yellow arrow regions). R, radius; U, ulna. **c**, Mice at E15.5 were labelled with BrdU for 2 h; labelled cells were detected using an anti-BrdU antibody in consecutive sections. **d**, Quantitative analysis of the labelled cells in *Ihh*^{E95K/E95K} compared to wild-type mice ($n = 3$; $*P = 0.036$, Student's *t*-test). Error bars are s.d. **e**, Interaction of the amino-terminal fragments of IHH (N-IHH) and IHH(E95K) (N-IHH(E95K)) with PTCH1. The binding affinity was determined using a cell-based competition assay for the binding of ³²P-labelled N-IHH protein to cells expressing PTCH1 (ref. 11), with an increasing concentration of unlabelled N-IHH or N-IHH(E95K). The K_d for PTCH1 determined from the competition graphs showed a clear difference between N-IHH and N-IHH(E95K) as indicated. The gel insert shows a stability test of the purified N-IHH and N-IHH(E95K) proteins under conditions of the binding assay assessed by western blot, indicating that there was no notable degradation of N-IHH or N-IHH(E95K) within the assay period. Scale bars, 1 mm (**a**) and 200 μ m (**b**, **c**).

of the future joint (Fig. 2a). Thus, HIP1 in the developing digit could assist other hedgehog partners in maintaining IHH signalling within range of the cartilage anlagen. *Ihh* is initially expressed in the centre of the condensing mesenchyme of the cartilage elements that flank the interzone regions from E13.5. Morphologically, these *Ihh*-expressing cells are no different from the surrounding cells (data not shown). The function of *Ihh* expression at this stage of digit development has not been well studied, but *Ptch1* and *Gli1* (Fig. 2a and Supplementary Fig. 5) expression patterns clearly indicate IHH activity throughout the condensing mesenchyme. IHH signalling is less intense in the interzone regions, where *Gdf5* is expressed (Fig. 2a). In the developing digits of BDA1 mice, there seemed to be a general decrease in IHH activity but, as in the growth plate, it extended further from the source than in

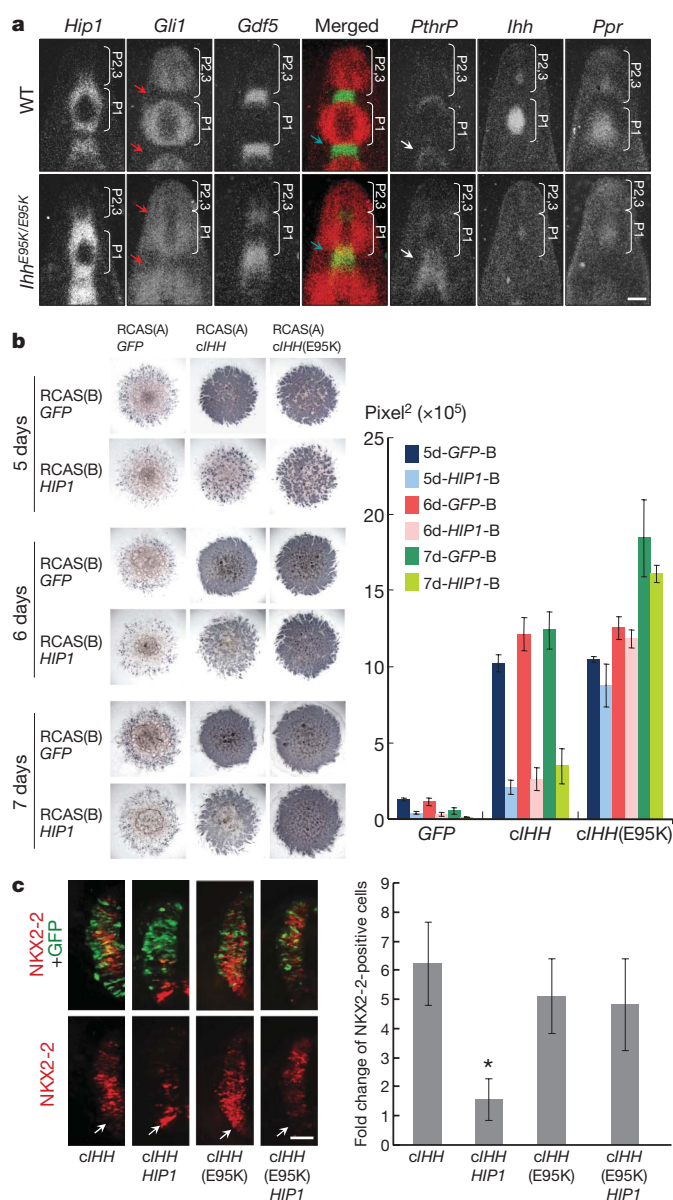


Figure 2 | IHH signals into the interzone in the developing digit of BDA1 mice owing to impaired interaction with HIP1. **a**, Molecular analysis of digit formation at E13.5. Expression of *Hip1*, *Gli1*, *Gdf5*, *PthrP*, *Ihh* and *Ppr* (also known as *Pth1r*) using radioactive *in situ* hybridization on sagittal sections of digit III. *Hip1* was expressed at the margin demarcating the cartilage element. *Gli1* (red) and *Gdf5* (green) expression signals from consecutive sections are merged, showing increased hedgehog signalling in the developing interzone (yellow; blue arrow region). Diffused *Gli1* expression was noted (red arrows), indicating wider spread of the hedgehog signal, and concomitant with an expanded interzone indicated by *Gdf5* expression. Increased *PthrP* expression in the developing interzone was also noted (white arrows). *Ppr* was expressed in the same domains as *Ihh*. P1, proximal phalange; P2, middle phalange; P3, distal phalange. **b**, Micromass cultures infected with RCAS(A) constructs and co-infected with RCAS(B) constructs carrying either GFP (control) or *HIP1*. Representative cultures stained for alkaline phosphatase (ALP) activity at indicated times are shown on the left; quantification of the stained surface area using the morphometric tool AutMess is shown on the right ($n = 4$). Chicken *IHH* and *IHH*(E95K) robustly induce ALP activity. Co-infection of RCAS(B)*HIP1* markedly reduces ALP induction by *cIHH*, but only mildly inhibits ALP induction by *cIHH*(E95K). **c**, Induction of NKX2-2 in response to overexpression of *cIHH* or *cIHH*(E95K) in the presence or absence of *HIP1* in the developing chick neural tube. The arrows indicate region of the endogenous *vp3* progenitor cells. The NKX2-2-expressing cells were counted (three sections per neural tube, $n = 5$), and the fold change in NKX2-2-positive cells relative to a control GFP electroporation is shown. $*P < 0.0001$, Student's *t*-test. Scale bars, 100 μ m (**a**) and 50 μ m (**c**); error bars are s.d.

wild-type mice, overlapping with the *Gdf5*-expressing domain of the interzone (Fig. 2a). This signalling pattern is in keeping with an increased movement of IHH(E95K) owing to impaired interaction with HIP1 and other partners. Consistent with an increased IHH signalling range, *Hip1*, *Gli1* and *PthrP* expression were increased in the interzone (Fig. 2a).

BDA1 mutations in IHH affect its binding to HIP⁵. We tested the consequence of the E95K mutation on hedgehog signalling using a chick limb bud micromass culture assay¹⁵. Overexpression of IHH impaired the chondrogenic differentiation of limb bud mesenchymal cells, instead promoting differentiation to cells expressing alkaline phosphatase (Fig. 2b). This effect of IHH was suppressed in the presence of HIP1, supporting HIP1 as a negative regulator of IHH signalling (Fig. 2b). IHH(E95K) similarly promoted differentiation, but this effect was not suppressed in the presence of HIP1 (Fig. 2b). Thus, the regulation of hedgehog signalling by HIP1 is impaired by the E95K mutation. A similar finding was observed when chondrogenesis was used as readout (Supplementary Fig. 6). We studied this further using *in ovo* electroporation to introduce IHH or IHH(E95K) into the developing chick neural tube¹⁶. In response to the endogenous SHH signal from the notochord and floor plate, *NKX2-2* is normally activated in ventral progenitor cells closest to the SHH source¹⁷. Forced expression of chicken IHH (cIHH) resulted in non-cell autonomous activation of *NKX2-2* in the electroporated region of the neural tube (Fig. 2c). This inducing activity was markedly reduced when HIP1 was co-expressed with IHH (Fig. 2c). Overexpression of cIHH(E95K) also induced *NKX2-2*, however, this effect was not inhibited by coexpression of HIP1. This is consistent with the *in vitro* finding.

Thus, in the growth plate, the impaired interaction of IHH(E95K) with its receptor PTCH1 may reduce the immediate signalling capacity of IHH emanating from the prehypertrophic chondrocytes. Reduced binding to PTCH1 and other partners such as HIP1 would allow IHH(E95K) to travel further, increasing its stimulation of *PthrP* transcription at the perichondrium. Together, these abnormalities cause reduced bone growth in the BDA1 mouse. Fine-tuning of the IHH signal may be important for regulating bone growth; indeed, recent genome-wide genetic association studies in two large populations of loci that influence adult height identified genes in the hedgehog signalling pathway, including *IHH*, *HIP1* and *PTCH1* (refs 18, 19).

For the digits, we propose that the enhanced signalling range of IHH(E95K) in BDA1 mouse is the primary stimulus underlying the defect. This is supported by findings that activation of hedgehog signalling in chondrocytes by conditional ablation of *Ptch1* leads to joint defects²⁰. In heterozygous *short digit* mice, an inversion in the *Shh* locus ectopically activates *Shh* in the developing phalanges in the proximal digital elements, resulting in increased *PthrP* in the interzone region, loss or reduced *Ihh* expression in distal digit elements, and mis-expression of *Gdf5*; a combination that results in a BDA1 phenotype²¹.

A question remains as to how the shortened or missing middle phalanges can be explained in the BDA1 mouse model. Measurement of the most distal cartilage element before the last joint that is formed showed this element to be markedly shorter in the BDA1 mouse (Fig. 3a). Using pulse-chase labelling with BrdU, we showed that in wild-type mice, proximodistal outgrowth involves the recruitment of surrounding proliferating mesenchymal cells into the condensing distal cartilage digit element (Fig. 3b), and that this process is reduced in the BDA1 mouse to about 40% of normal (Fig. 3c, d). Although FGF signalling from the apical ectodermal ridge is thought to be important in regulating the size of the cartilage elements, which can influence the location and number of phalangeal segmentations²², *Fgf8* expression was normal in BDA1 mice (Supplementary Fig. 7), suggesting that IHH may have a more direct role. Indeed, we showed intense *Gli1* expression at the distal tip of wild-type mice from E13.5, which was reduced in the BDA1 mouse (Fig. 3e), consistent with reduced *Ihh* expression in the most distal cartilage element in the BDA1 mouse (Fig. 3f). Therefore, in addition

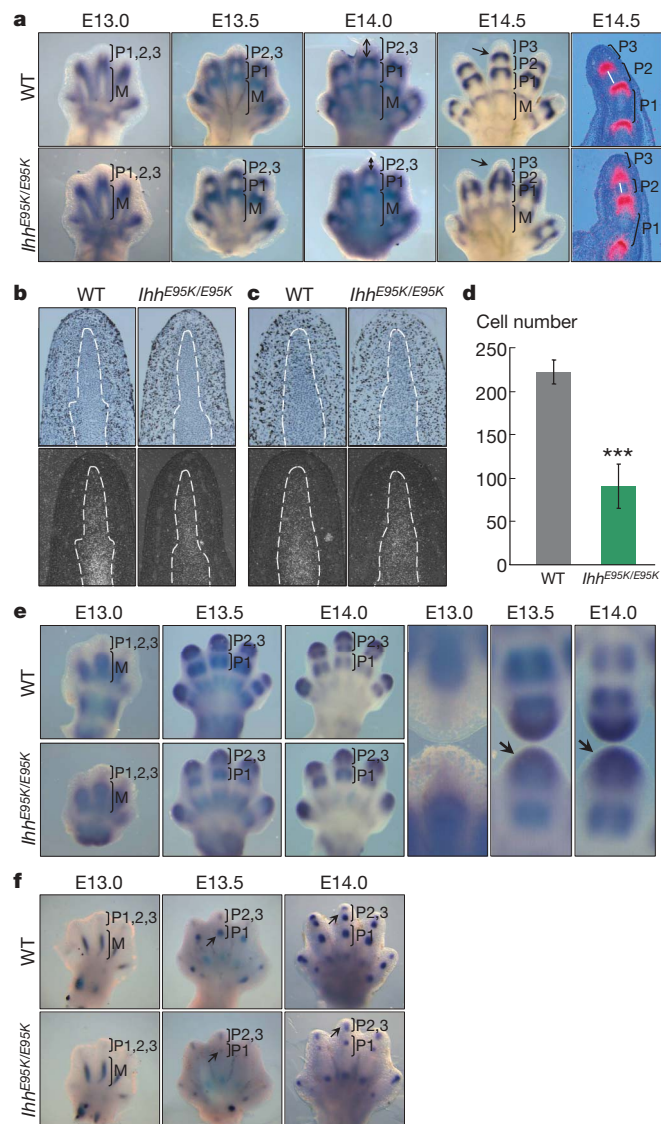


Figure 3 | Impaired digit outgrowth in BDA1 mouse from reduced IHH

signalling at the distal tip. **a**, *Gdf5* expression in the developing digits of the hind limb from E13.0 to E14.5. Cartilage elements in digit III (M, metatarsal; P1, proximal phalange; P2, middle phalange; P3, distal phalange) showed no notable developmental difference between wild-type (WT) and *Ihh*^{E95K/E95K} mice at E13.0. From E13.5, a slight reduction in distal digit outgrowth was noted that became obvious by E14.0 (regions marked by double-arrow lines), with shortened P2 and P3 elements at E14.5. Radioactive *in situ* hybridization of sagittal sections for digit III at E14.5 highlights the shortened middle phalangeal element (white lines). **b**, Two-hour pulse-labelling with BrdU. BrdU staining showed proliferating mesenchymal cells surrounding the chondrogenic condensation (regions enclosed by white dotted lines) delineated by the expression of *Col2a1* in a consecutive section (shown below the corresponding sections). No difference was found between wild-type and *Ihh*^{E95K/E95K} mice. **c**, Two-hour pulse/10-h chase-labelling with BrdU. After a 10-h chase, BrdU-positive cells were detected within the distal *Col2a1* expression domain, which is fewer in *Ihh*^{E95K/E95K} than wild-type mice, indicating impaired recruitment at the distal tip region. **d**, Quantitative analysis of BrdU-positive cells showing a significant decrease in *Ihh*^{E95K/E95K} mice ($n = 5$; error bars indicate s.d.; *** $P < 0.0001$, Student's *t*-test). **e**, Whole-mount *in situ* hybridization for the expression of *Gli1* in hind limb digits from E13.0 to E14.0, showing reduced expression at the tip region in *Ihh*^{E95K/E95K} mice. To demonstrate the reduction, digits III of the wild-type and *Ihh*^{E95K/E95K} mice are compared tip-to-tip with higher magnification in panels to the right (black arrow regions). **f**, Expression of *Ihh* in digit cartilage elements of *Ihh*^{E95K/E95K} mice was similar to wild-type at E13.0, but reduced at E13.5 and E14.0 (black arrow). The separation of P2 and P3 *Ihh* expression domains was delayed in *Ihh*^{E95K/E95K} mice at E14.0.

to FGF, IHH signal from the condensed mesenchyme is also important for proper distal digit outgrowth. Recently, SHH was shown to function in a growth-promoting phase in the recruitment of digit precursor mesenchyme before E13.0 (ref. 8). Our data indicate IHH functions similarly, but acts later (after E13.0) in the outgrowth of distal phalangeal elements.

Our findings provide, to our knowledge, the first insight into the *in vivo* consequence of hedgehog mutations that impair interaction with hedgehog partners, and strongly implicate altered IHH signalling capacity and range in the pathogenesis of BDA1 (Fig. 4a). Our model for enhanced long-range signalling is consistent with a theoretical model in which an increase in the effective diffusion coefficient, such as that caused by reducing retention on binding proteins, lowers the

responses close to the source but increases those at a distance²³. We propose a mechanism for the digit abnormality in BDA1 (Fig. 4b, c) in which reduced IHH signalling at the digit tip region results in shortened cartilage elements that are either too short for the last joint to be formed (digit V in the BDA1 mouse), or form a joint with severely shortened middle phalanges (digits II–IV in the BDA1 mouse). In mice, this phenotype is most prominent in the homozygous state, whereas BDA1 is dominant in humans. The reason for this is not known but it is not uncommon^{24,25}, and it may be related to modifying genes and the variations in the capacity of a signalling molecule to induce an effect between mouse and human.

METHODS SUMMARY

Embryonic stem cells carrying the E95K mutation were generated by homologous recombination. These cells were used to produce mouse chimaeras that were crossed with β -actin *Cre* mice²⁶ to remove the neomycin (*neo*) cassette in conjunction with germline transmission of the targeted allele. Mice of 129SvEv or mixed 129SvEv-ICR backgrounds were used as there were no differences in the phenotypic outcome. X-rays were taken using a General Electric X-ray machine (Senograph 600T Senix HF, General Electric). Skeletal preparations were stained with Alizarin Red S and Alcian Blue²⁷. Pulse-chase labelling with BrdU was performed by injection of thymidine to dilute BrdU incorporation during the chase period²⁸. *In situ* hybridization and immunostaining were performed on 5- μ m dewaxed sections using [³⁵S]-UTP-labelled riboprobes²⁹. *IHH* and *IHH*^{E95K}, and *HIP1* were overexpressed in chick limb bud micromass cultures using the RCAS(A) and RCAS(B) viruses, respectively¹⁵. These genes were also overexpressed in the developing chick neural tube by electroporation of the respective expression constructs¹⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 July 2008; accepted 12 February 2009.

Published online 1 March 2009.

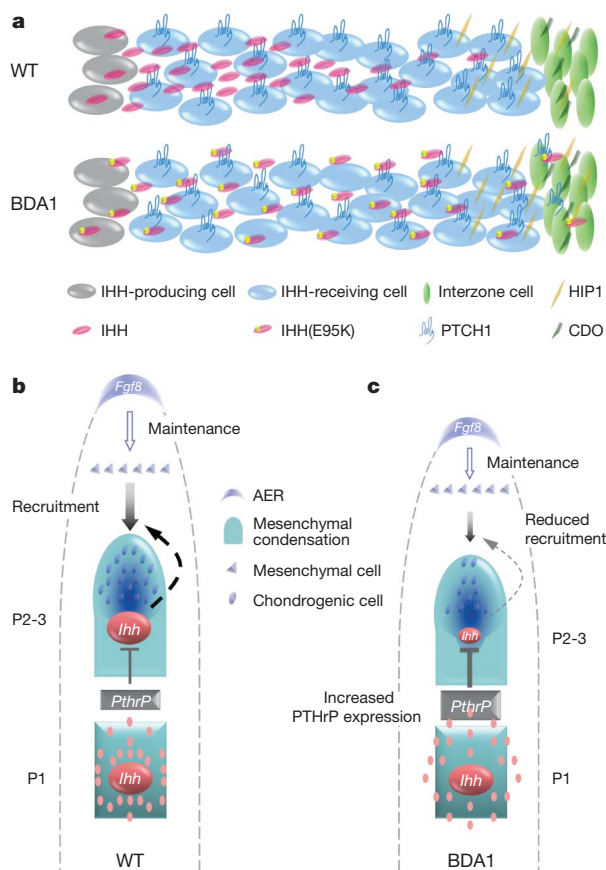


Figure 4 | Model for the molecular mechanism of the E95K mutation altering IHH signalling causing BDA1. **a**, Model illustrating the normal range of IHH signalling in the developing digit in which interaction with PTCH1 and HIP1 retains the signal within the cartilage element. In the BDA1 mouse the interaction with PTCH1 and HIP1 is weakened, allowing the IHH(E95K) to diffuse further into the interzone region. **b**, The proposed role of IHH during early digit patterning and its regulation. *Ihh* expressed in the distal phalangeal condensation signals to the undifferentiated proliferating mesenchymal cells under the apical ectodermal ridge and promotes recruitment/chondrogenesis for digit outgrowth in conjunction with FGF signalling. We propose that *Ihh* expression in the distal elements is regulated by signals mediated in the proximal interzone, probably PTHrP as *Ppr* is expressed in the *Ihh* domain, through the negative feedback loop modulating chondrocyte differentiation. **c**, In BDA1 mice, the range of IHH signal is increased in the proximal element, extending further into the developing interzone region, upregulating genes such as *PthrP*. This increases its signal to the distal condensed mesenchyme where *Ppr* is expressed and downregulates *Ihh* expression, reducing recruitment/chondrogenesis and thereby the size of the distal element. Thus, P2–P3 segmentation begins within a smaller chondrogenic template, resulting in a smaller template for the future middle phalanx (P2). P1, proximal phalanx.

- Gao, B. *et al.* Mutations in *IHH*, encoding Indian hedgehog, cause brachydactyly type A-1. *Nature Genet.* **28**, 386–388 (2001).
- McCready, M. E., Grimsey, A., Styer, T., Nikkel, S. M. & Bulman, D. E. A century later Farabee has his mutation. *Hum. Genet.* **117**, 285–287 (2005).
- Hooper, J. E. & Scott, M. P. Communicating with Hedgehogs. *Nature Rev. Mol. Cell Biol.* **6**, 306–317 (2005).
- Jiang, J. & Hui, C. C. Hedgehog signaling in development and cancer. *Dev. Cell* **15**, 801–812 (2008).
- McLellan, J. S. *et al.* The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. *Nature* **455**, 979–983 (2008).
- Wallis, D. & Muenke, M. Mutations in holoprosencephaly. *Hum. Mutat.* **16**, 99–108 (2000).
- Hellems, J. *et al.* Homozygous mutations in *IHH* cause acrocapitofemoral dysplasia, an autosomal recessive disorder with cone-shaped epiphyses in hands and hips. *Am. J. Hum. Genet.* **72**, 1040–1046 (2003).
- Zhu, J. *et al.* Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. *Dev. Cell* **14**, 624–632 (2008).
- Kronenberg, H. M. Developmental regulation of the growth plate. *Nature* **423**, 332–336 (2003).
- St-Jacques, B., Hammerschmidt, M. & McMahon, A. P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev.* **13**, 2072–2086 (1999); erratum **13**, 2617 (1999).
- Fuse, N. *et al.* Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proc. Natl Acad. Sci. USA* **96**, 10992–10999 (1999).
- Vortkamp, A. *et al.* Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* **273**, 613–622 (1996).
- Tenzen, T. *et al.* The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev. Cell* **10**, 647–656 (2006).
- Chuang, P. T. & McMahon, A. P. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* **397**, 617–621 (1999).
- Seemann, P. *et al.* Activating and deactivating mutations in the receptor interaction site of GDF5 cause symphalangism or brachydactyly type A2. *J. Clin. Invest.* **115**, 2373–2381 (2005).
- Cheung, M. & Briscoe, J. Neural crest development is regulated by the transcription factor Sox9. *Development* **130**, 5681–5693 (2003).
- Ericson, J. *et al.* Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell* **90**, 169–180 (1997).
- Lettre, G. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nature Genet.* **40**, 584–591 (2008).
- Weedon, M. N. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genet.* **40**, 575–583 (2008).

20. Mak, K. K., Chen, M. H., Day, T. F., Chuang, P. T. & Yang, Y. Wnt/ β -catenin signaling interacts differentially with Ihh signaling in controlling endochondral bone and synovial joint formation. *Development* **133**, 3695–3707 (2006).
21. Niedermaier, M. *et al.* An inversion involving the mouse *Shh* locus results in brachydactyly through dysregulation of *Shh* expression. *J. Clin. Invest.* **115**, 900–909 (2005).
22. Sanz-Ezquerro, J. J. & Tickle, C. Fgf signaling controls the number of phalanges and tip formation in developing digits. *Curr. Biol.* **13**, 1830–1836 (2003).
23. Saha, K. & Schaffer, D. V. Signal dynamics in Sonic hedgehog tissue patterning. *Development* **133**, 889–900 (2006).
24. Gong, Y. *et al.* Heterozygous mutations in the gene encoding noggin affect human joint morphogenesis. *Nature Genet.* **21**, 302–304 (1999).
25. Polinkovsky, A. *et al.* Mutations in *CDMP1* cause autosomal dominant brachydactyly type C. *Nature Genet.* **17**, 18–19 (1997).
26. Meyers, E. N., Lewandoski, M. & Martin, G. R. An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nature Genet.* **18**, 136–141 (1998).
27. McLeod, M. J. Differential staining of cartilage and bone in whole mouse fetuses by alcian blue and alizarin red S. *Teratology* **22**, 299–301 (1980).
28. Tsang, K. Y. *et al.* Surviving endoplasmic reticulum stress is coupled to altered chondrocyte differentiation and function. *PLoS Biol.* **5**, e44 (2007).
29. Wai, A. W. *et al.* Disrupted expression of matrix genes in the growth plate of the mouse cartilage matrix deficiency (*cmd*) mutant. *Dev. Genet.* **22**, 349–358 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the Research Grants Council and University Grants Council of Hong Kong (N_HKU705/02, HKU2/02C and AoE/M-04/04), and The National Key Scientific Program (2007CB947300). We thank P. Tam for suggestions and comments, P. Beachy and D. Leahy for sharing unpublished data, and K. Leung for blastocyst injections.

Author Contributions The primary affiliation for J.H. is Bio-X Center, Shanghai Jiao Tong University. B.G. and J.H. designed and performed the experiments and analysed the data. K.F.L. performed the *in vitro* hedgehog signalling assay. F.W., S.S. and S.M. were involved in the *in situ* analysis of digit formation, micromass cultures and revision of the manuscript. M.C. and J.B. assisted in the chick neural tube electroporation experiments. G.M. performed the PTCH1 competitive binding assay. L.H. was involved in experimental design and manuscript revision. K.S.E.C. was involved in experimental design, generation of the BDA1 mouse, data analysis, interpretation and manuscript revision. D.C. coordinated the experimental design, analysed the data, and together with B.G. and J.H. interpreted the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.C. (chand@hkusua.hku.hk) or L.H. (helinhelin@gmail.com).

METHODS

Generation of gene-targeted BDA1 mice. A targeting construct containing a 397G > A substitution (p.E95K mutation)¹ was generated in a 6.2-kb Sca-I/Stu-I *Ihh* genomic DNA fragment (gift from A. P. McMahon), together with a *neo* cassette flanked with *loxP* sites inserted in intron 1 (Supplementary Fig. 1a). Gene targeting was performed in L4 embryonic stem cells, derived from 129Sv/ev blastocysts (K. Cheah, unpublished data). G418-resistant clones with the appropriate homologous recombination events were determined from Southern blot analysis of Eco-RI digested DNA, using internal (3') or external (5' and 3') probes (Supplementary Fig. 1a, b). Mouse chimaeras were generated from targeted embryonic stem cells by blastocyst injections³⁰, and mice heterozygous for the E95K allele were generated from the chimaeras crossed with β -actin Cre mice²⁶ to remove the *neo* cassette. C57BL/6 mice heterozygous for an *Ihh*-null allele (*Ihh*^{+/-}) were obtained from A. P. McMahon¹⁰.

Radiological analysis. X-rays were taken using Kodak Diagnostic film Min-R H MRH-1 on a Senograph 600T Senix HF machine (General Electric). Limb bone lengths were measured on the radiographs, and statistical analysis was performed using Student's *t*-test, with a two-tailed *P* < 0.05 taken as statistically significant.

Skeletal staining. E13.5–E18.5 fetuses, newborn and P10 mice were eviscerated, fixed overnight in 95% ethanol, and then transferred sequentially to 100% ethanol and 100% acetone, followed by staining with Alizarin Red S and Alcian Blue²⁷.

In situ hybridization. *In situ* hybridization on dewaxed sections was performed as previously described²⁹, using [³⁵S]-UTP- or digoxigenin-labelled riboprobes for *Col10a1* and *Col2a1* (ref. 29), *PTHrP* and *Ppr*³¹, *Ihh*³², *Gli1*, *Gdf5*, *Ptch1* (ref. 33), *Wnt9a* (ref. 34) and *Fgf8* (ref. 35). For whole-mount *in situ* hybridization, anti-digoxigenin antibody conjugated with alkaline phosphatase (Roche) and Boehringer Mannheim purple substrate (Roche) were used to visualize the signal. All markers were analysed on sections from at least three wild-type and mutant mice.

BrdU incorporation. For pulse labelling of fetal mice, pregnant females at a specified gestational stage were injected intraperitoneally with 200 μ g of BrdU per gram of body weight and killed after 2 h. For pulse-chase labelling, after the initial 2 h incorporation, 6 mg of thymidine (Sigma) was injected to dilute the BrdU, stopping further incorporation^{36,37}, and the mice were killed 10 h later. BrdU-positive cells were identified by immunostaining using a staining kit (Zymed). Expression of *Col2a1* (*in situ* hybridization) on a consecutive section was used to determine the region of the chondrogenic condensation, and the number of BrdU-positive cells in the *Col2a1* region was calculated on several sections from five wild-type and five *Ihh*^{E95K/E95K} mice.

Luciferase hedgehog signalling assay. A hedgehog-signalling reporter cell line was produced in C3H10T1/2 cells stably co-transfected with an 8 \times GliBS-luc reporter DNA plasmid (gift from P. Beachy), a CMV-lacZ cassette contained in a 3.8 kb Mun-I/Xho-I fragment excised from pcDNA3.1/V5-His/lacZ (Invitrogen), and a 6.0 kb Sal-I and Eco-RI pGK-neo cassette³⁸ for colony selection. Hedgehog signalling (luciferase activity) was measured as described with modifications³⁹. Reporter cells in 24-well plates were cultured in conditioned medium from COS-7 cells stably expressing IHH and IHH(E95K). After 48 h, luciferase activity from the cells was determined using a kit from Promega, normalized to the concentration of wild-type and IHH(E95K), and to cell number using β -galactosidase activity (Luminescent β -galactosidase reporter system 3; BD Biosciences) as an internal control. Statistical analysis was performed using a Student's *t*-test.

Cell-based assay for the binding between PTCH1 and IHH. Binding affinity of the functional N-terminal fragment of IHH (N-IHH) and IHH(E95K) (N-IHH(E95K)) to PTCH1 was evaluated using an established assay¹¹. In brief, glutathione *S*-transferase (GST)-tagged N-IHH and N-IHH(E95K) (pGEX-2T vector) expressed in *Escherichia coli* were affinity purified and the GST tag was removed by thrombin digestion. N-IHH proteins were enzymatically labelled using [γ -³²P]-ATP at a protein kinase A (PKA) site engineered at the carboxy terminus. Assays were carried out using ECR-CHO cells stably transfected for ecdysone-inducible expression of Ptc-CTD140 (ref. 11). After induction with ponasterone A, 3 \times 10⁵ activated cells were incubated with 20 nM of [³²P]-N-IHH and different concentrations of unlabelled wild-type N-IHH or

N-IHH(E95K) at 4 °C for 2 h. Binding of [³²P]-N-IHH in the presence of N-IHH or N-IHH(E95K) was normalized to the total amount of bound [³²P]-N-IHH in the absence of competitors set at 100%, and the N-IHH/PTCH1-CTD140 binding affinity determined from a plot of the bound [³²P]-N-IHH against an increasing concentration of unlabelled N-IHH and N-IHH(E95K) proteins.

Chick micromass cultures. RCAS(A)-cIHH carrying the chicken *IHH* was provided by A. Vortkamp. RCAS(A)-cIHH(E95K) was constructed by site-directed mutagenesis (QuikChange kit, Stratagene). Viral supernatant was produced as previously described⁴⁰. Chicken micromass cultures¹⁵ were infected with RCAS(A)-GFP, RCAS(A)-cIHH or RCAS(A)-cIHH(E95K). Co-infection of micromass cultures was performed with either RCAS(B)-GFP as a control or RCAS(B)-HIP. Cultures were incubated for 5–7 days and stained with Alcian blue for cartilaginous condensations or with alkaline phosphatase for differentiated chondrocytes, and quantified using the Zeiss AutMess analytics tool.

Electroporation and expression of genes in the developing chick neural tube. Fertilized white Leghorn eggs were obtained from Tin Hang Technology Limited and incubated in a humidified incubator at 38 °C. Embryos were staged as described previously⁴¹. *In ovo* electroporation was carried out as described¹⁶. Expression constructs were generated in the pCAGGS vector⁴² for chicken *IHH*, *IHH*^{E95K} and *HIP1*, then inserted upstream of an internal ribosomal entry site and nuclear localization sequence tagged-GFP. For electroporation, plasmid DNA was injected into the lumen of Hamburger Hamilton stage 11–12 neural tubes. Electroporation was carried out using a BTX electroporator delivering five 50-ms pulses of 30V across the neural tube. Transfected embryos were incubated for 24 h before processing for immunohistochemistry using antibodies for GFP (AB-Direct) and NKX2-2 (DSHB) as described⁴³. The monoclonal antibodies against NKX2-2 (74.5A5, developed by T. M. Jessell) were obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa Department of Biological Sciences.

30. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. R. *Manipulating the Mouse Embryo. A Laboratory Manual* (Cold Spring Harbour Laboratory Press, 2003).
31. Lee, K. *et al.* Parathyroid hormone-related peptide delays terminal differentiation of chondrocytes during endochondral bone development. *Endocrinology* **137**, 5109–5118 (1996).
32. Bitgood, M. J. & McMahon, A. P. *Hedgehog* and *Bmp* genes are coexpressed at many diverse sites of cell–cell interaction in the mouse embryo. *Dev. Biol.* **172**, 126–138 (1995).
33. Mundlos, S. Skeletal morphogenesis. *Methods Mol. Biol.* **136**, 61–70 (2000).
34. Spater, D. *et al.* Wnt9a signaling is required for joint integrity and regulation of *Ihh* during chondrogenesis. *Development* **133**, 3039–3049 (2006).
35. Crossley, P. H. & Martin, G. R. The mouse *Fgf8* gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439–451 (1995).
36. Haaf, T. High-resolution analysis of DNA replication in released chromatin fibers containing 5-bromodeoxyuridine. *Biotechniques* **21**, 1050–1054 (1996).
37. Schmahl, J., Eicher, E. M., Washburn, L. L. & Capel, B. Sry induces cell proliferation in the mouse gonad. *Development* **127**, 65–73 (2000).
38. Adra, C. N., Boer, P. H. & McBurney, M. W. Cloning and expression of the mouse *pgk-1* gene and the nucleotide sequence of its promoter. *Gene* **60**, 65–74 (1987).
39. Taipale, J. *et al.* Effects of oncogenic mutations in *Smoothed* and *Patched* can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).
40. Logan, M. & Tabin, C. Targeted gene misexpression in chick limb buds using avian replication-competent retroviruses. *Methods* **14**, 407–420 (1998).
41. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. 1951. *Dev. Dyn.* **195**, 231–272 (1992).
42. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–199 (1991).
43. Briscoe, J., Chen, Y., Jessell, T. M. & Struhl, G. A hedgehog-insensitive form of *Patched* provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol. Cell* **7**, 1279–1291 (2001).

Embryonic stem cells use ZFP809 to silence retroviral DNAs

Daniel Wolf¹ & Stephen P. Goff¹

Embryonic stem cells (ESCs) and other primitive stem cells of mice have been known for more than 30 years to potentially block retrovirus replication¹. Infection of ESCs by the murine leukaemia viruses (MLVs) results in the normal establishment of integrated proviral DNA, but this DNA is then transcriptionally silenced, preventing further viral spread. The repression is largely mediated by *trans*-acting factors that recognize a conserved sequence element termed the primer binding site, an 18-base pair sequence complementary to the 3' end of a cellular transfer RNA^{2–6}. A specific tRNA is annealed to the primer binding site sequence of the viral genomic RNA, and is used to prime DNA synthesis⁷. This same sequence in the context of the integrated proviral DNA is targeted for silencing in ESCs. We have recently shown that a large protein complex binding to the primer binding site in ESCs contains TRIM28 (refs 8, 9), a well-characterized transcriptional co-repressor^{10–12}. An important question remains as to the identity of the factor that directly recognizes integrated retroviral DNAs and recruits TRIM28 to mediate their specific silencing. Here we identify the zinc finger protein ZFP809 as the recognition molecule that bridges the integrated proviral DNA and TRIM28. We show that expression of ZFP809 is sufficient to render even differentiated cells highly resistant to MLV infection. Furthermore, we demonstrate that ZFP809 is able to potentially block transcription from DNA constructs of human T-cell lymphotropic virus-1 (HTLV-1), which use the same primer tRNA. These results identify ZFP809 as a DNA-binding factor that specifically recognizes a large subset of mammalian retroviruses and retroelements, targeting them for transcriptional silencing. We propose that ZFP809 evolved as a stem-cell-specific retroviral restriction factor, and therefore constitutes a new component of the intrinsic immune system of stem cells.

Moloney MLV, a prototypical simple retrovirus, contains a primer binding site (PBS) complementary to proline tRNA (PBS^{Pro}), and is potentially restricted by ESCs and embryonic carcinoma cells (ECCs) in a PBS-dependent mechanism. A single base-pair change known as the B2 mutation (PBS^{B2}) is sufficient to prevent binding of the ESC silencing factors and so abrogate this restriction². To identify the DNA-binding component of the complex, we used an ultraviolet-crosslinking approach. We first confirmed that a BrdU-substituted radiolabelled DNA probe corresponding to PBS^{Pro} retained the correct specificity for the PBS silencing complex (Supplementary Fig. 1). We then incubated this DNA probe with nuclear extracts of the F9 ECC cell line, treated the samples with ultraviolet light to form covalent protein–DNA crosslinks, and resolved the complexes by SDS–PAGE under denaturing conditions. A crosslinked species migrating at the position of an approximately 45-kDa protein was observed. This species was not detected when the experiment was performed using a mutant PBS^{B2} probe, or when a large excess of unlabelled (cold) PBS^{Pro} probe was added to the reaction (Fig. 1a). The

addition of 15 base pairs (bp) to the PBS^{Pro} probe caused the species to migrate at a position corresponding to a mass of ~55 kDa. To confirm that the 45-kDa species was a true component of the PBS binding complex, we showed that this species co-immunoprecipitated with TRIM28, a confirmed subunit of the silencing complex (Fig. 1b). These experiments suggested that a 45-kDa protein was in close contact with DNA and thus was a strong candidate for the sequence-specific recognition factor.

To determine the identity of this protein, we partially purified the PBS silencing complex from F9 ECC nuclear extracts⁸ and immunoprecipitated the complex with a TRIM28-specific antibody (Fig. 1c). This immunoprecipitated material was resolved by SDS–PAGE, and bands corresponding to the 45-kDa species were excised and characterized by liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) (Fig. 1d and Supplementary Table 1). Of the proteins identified, the three most likely candidates for DNA-binding factors were L1TD1, ZFP809 and ZFP124. L1TD1 has sequence similarity to the products of the *ORF1* gene of LINE-1 retrotransposons, proteins known to bind DNA¹³. ZFP809 and ZFP124 are both members of the family of Kruppel-associated box (KRAB)-containing zinc finger proteins, which bind DNA¹⁴ and interact with TRIM28 via their KRAB domain^{15,16}. To determine which of these candidates has the requisite DNA-binding specificity, we expressed all three in *Escherichia coli* as recombinant proteins fused to glutathione-S-transferase (GST)¹⁷, and tested lysates in electrophoretic mobility shift assays (EMSAs) with 28-bp probes corresponding to either PBS^{Pro} or PBS^{B2} (Fig. 2a and Supplementary Fig. 2). Of the three candidates only ZFP809 bound to the wild-type PBS^{Pro} sequence (but not to PBS^{B2}), strongly suggesting that ZFP809 was the DNA recognition component of the PBS silencing complex (Fig. 2a). To determine whether ZFP809 was able to bind simultaneously to PBS^{Pro} DNA and TRIM28, we added increasing amounts of recombinant GST-tagged TRIM28 to EMSA reactions containing ZFP809 and either the PBS^{Pro} or the PBS^{B2} probe (Fig. 2b). The addition of TRIM28 markedly supershifted the ZFP809–PBS^{Pro} probe complex, whereas the PBS^{B2} probe remained unbound (Fig. 2b). The addition of TRIM28 markedly increased the yield of shifted probe, suggesting that it stabilized the complex or increased the affinity of ZFP809 for the PBS^{Pro} probe. Furthermore, the PBS binding complex that formed in the presence of recombinant TRIM28 and ZFP809 proteins showed virtually identical migration to the PBS complex formed in nuclear extracts from ECCs, suggesting that these two components alone are sufficient to reconstitute the minimal PBS silencing complex (Fig. 2b). Moreover, the DNA-binding specificity of this recombinant reconstituted complex was found to be virtually indistinguishable from the native complex (Supplementary Fig. 3).

These experiments strongly suggested that ZFP809 was the DNA recognition component of the PBS silencing complex and, because TRIM28 is ubiquitous and present at high concentrations, also indicated

¹Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, College of Physicians and Surgeons, HHSC 1310, 701 West 168th Street, New York, New York 10032, USA.

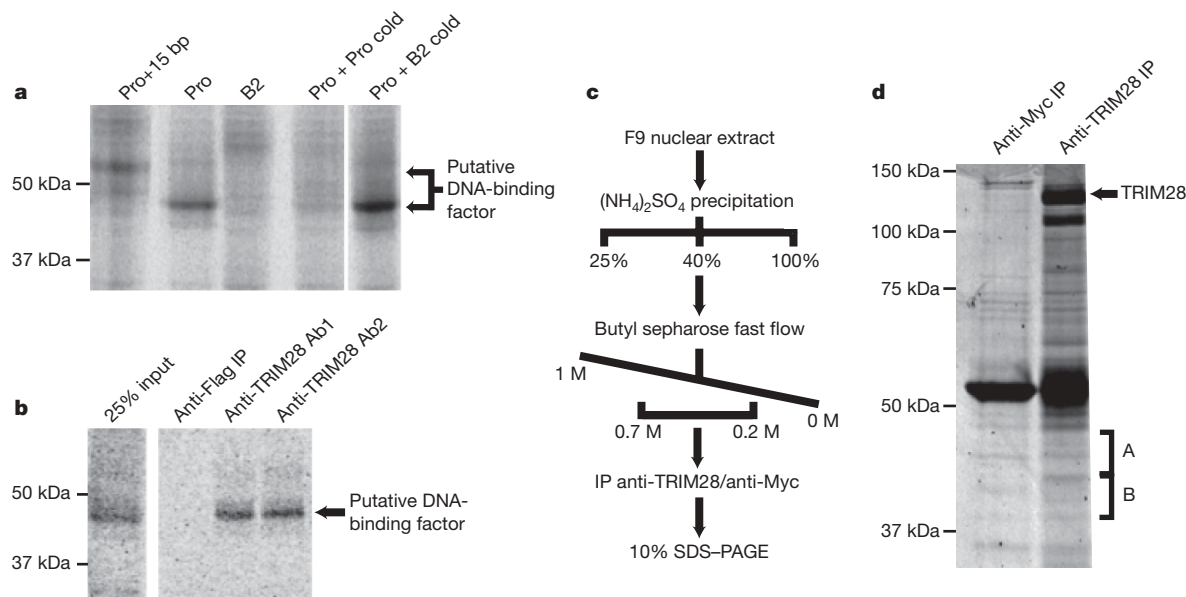


Figure 1 | Identification of candidate PBS^{Pro} DNA-binding factors. **a**, An F9 ECC nuclear extract was prepared and incubated with radiolabelled probes corresponding to PBS^{Pro} (Pro), PBS^{B2} (B2) or PBS^{Pro} with a 15-bp extension (Pro+15 bp). Crosslinking was performed by exposure to ultraviolet light. Competition was achieved by adding 100 pmol of unlabelled (cold) probe corresponding to either PBS^{Pro} or PBS^{B2}. **b**, Crosslinked nuclear extracts

were subjected to immunoprecipitation (IP) with anti-Flag or two separate anti-TRIM28 antibodies (Ab1 and Ab2)*. **c**, The purification scheme used to prepare samples for identification by mass spectrometry. **d**, SDS-PAGE of samples prepared by the purification scheme outlined in **c**, visualized by Coomassie stain. Bands excised for protein identification by mass spectrometry are labelled 'A' and 'B'.

that ZFP809 was probably the critical limiting component in ESCs and thus sufficient to induce virus resistance. We therefore addressed whether expression of this protein in a permissive cell line could render it resistant to MLV or other viruses using a PBS specific for proline tRNA. Initial attempts to generate 293A cell lines stably expressing full-length ZFP809 were unsuccessful and most often yielded clones with 3' deletions, suggesting that the carboxy-terminal 50 residues are toxic in non-stem cells. Subsequent experiments were performed with constructs expressing the ZFP809 protein truncated at amino acid 353 (ZFP809(1–353)), which showed minimal toxicity and was expected to retain its silencing functionality as it included both DNA-binding and TRIM28-binding domains (Supplementary Fig. 4). 293A cells stably expressing ZFP809(1–353) with an amino-terminal Flag tag were generated and individual clones were screened for ZFP809(1–353) expression (Supplementary Fig. 5b). Clones expressing the empty vector were constructed in parallel. Three representative cloned lines were selected for further study: clone 5, which expressed representative levels of ZFP809(1–353); clone 9, which had been transformed with the ZFP809(1–353) DNA but failed to express ZFP809; and control clone 1, which contained the empty vector (Fig. 3a and Supplementary Fig. 5b). The cell lines were all confirmed to retain normal TRIM28 expression (Fig. 3a). Nuclear extracts were prepared and tested for the presence of the silencing complex by EMSAs. The line expressing ZFP809(1–353) exhibited a strong PBS^{Pro} binding activity (Fig. 3b). This reconstituted PBS silencing complex showed the same specificity as the native ECC complex and did not bind to the PBS^{B2} probe (Fig. 3b). The shifted complex produced by expression of ZFP809(1–353) in 293A cells was shown to contain both endogenous TRIM28 and exogenous Flag-tagged ZFP809(1–353), as judged by supershifts with anti-TRIM28 and anti-Flag antibodies. The interaction between ZFP809 and TRIM28 *in vivo* was confirmed by co-immunoprecipitation in clone 5 cells (Fig. 3c).

To test whether reconstitution of the minimal PBS silencing complex was sufficient to induce silencing of MLV in a PBS-dependent manner, we used two MLV reporter constructs that were identical except for a single base-pair change in the PBS sequence. MLV particles pseudotyped by the VSV G protein, transducing the puromycin-resistance gene, and using either PBS^{Pro} or PBS^{B2}, were generated. The infectivity of these two virus preparations was determined by

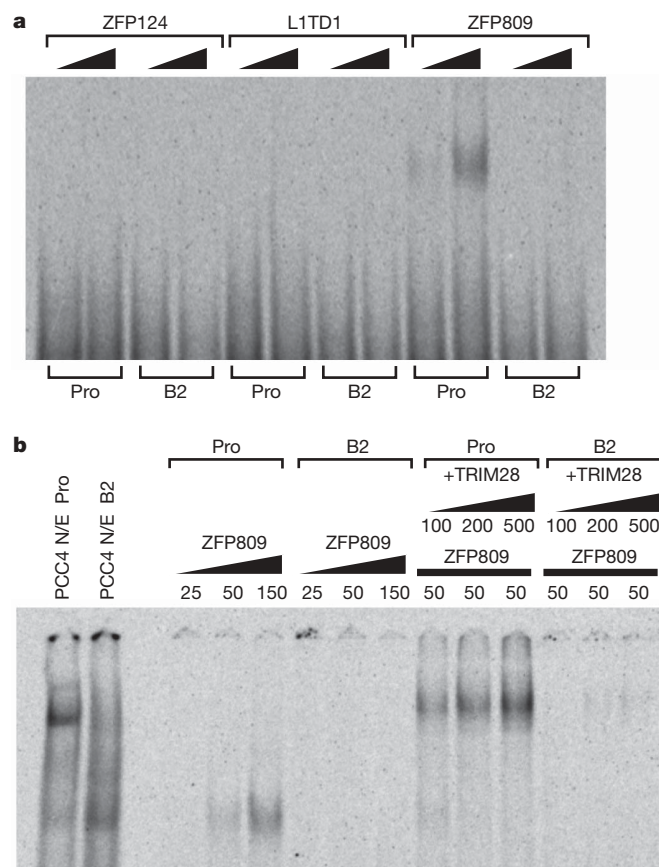


Figure 2 | Reconstitution of the PBS^{Pro} binding complex with recombinant proteins. **a**, Fifty or two-hundred-and-fifty nanograms of recombinant GST-tagged ZFP124, L1TD1 and ZFP809 were added to EMSA reactions with radiolabelled probes corresponding to PBS^{Pro} (Pro) or PBS^{B2} (B2). **b**, EMSAs performed as above, with PCC4 ECC nuclear extract (N/E), or with 25, 50 or 150 ng of GST-ZFP809 protein and 100, 200 or 500 ng of GST-TRIM28 protein, as indicated.

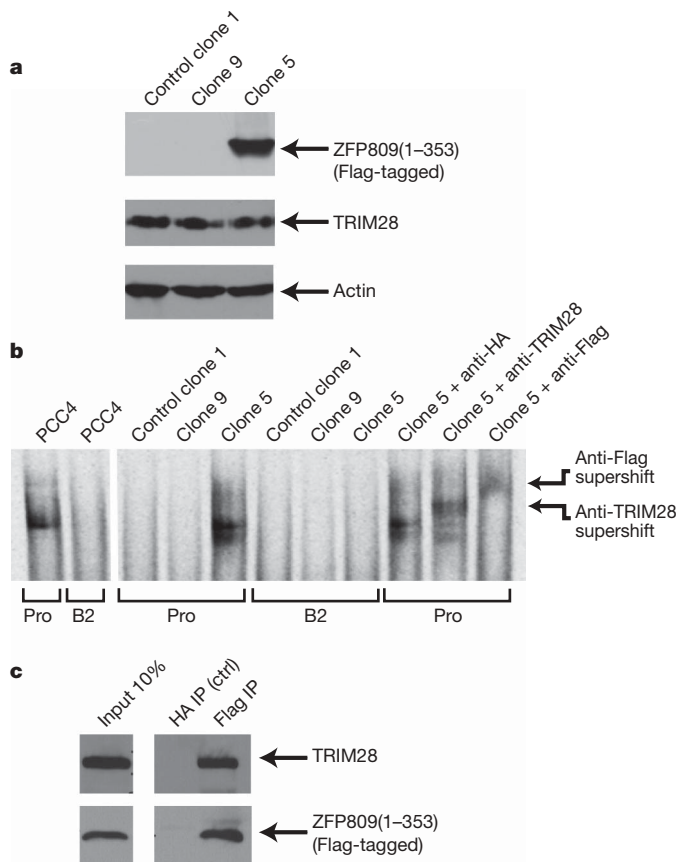


Figure 3 | Stable expression of ZFP809 in 293A cells leads to reconstitution of the PBS silencing complex in vivo. **a**, Nuclear extracts from clonal cell lines control clone 1 (empty vector control), clone 5 (cell line stably expressing Flag-tagged ZFP809(1–353)) or clone 9 (cell line generated with same construct as clone 5 but does not express ZFP809), probed by western blot with antisera for Flag (top panel), TRIM28 (middle panel) or β -actin (bottom panel). **b**, Nuclear extracts prepared from PCC4 ECCs or from cell lines control clone 1, clone 9 or clone 5 used in EMSA reactions with radiolabelled probes corresponding to PBS^{Pro} (Pro) or PBS^{B2} (B2). In the final three lanes the PBS silencing complex is supershifted by the addition of 2 μ g anti-TRIM28 (ref. 8), anti-Flag, or a control (ctrl) anti-haemagglutinin (HA) antibody. **c**, ZFP809(1–353) and TRIM28 co-immunoprecipitate in DNase-treated clone 5 whole cell extracts in the presence of ethidium bromide. Control immunoprecipitations were performed with HA antisera. Immunoprecipitated ZFP809(1–353) and co-immunoprecipitated TRIM28 were detected by western blot.

colony formation assays after infection of clone 5, and the controls consisting of control clone 1, clone 9 and the parental 293A cell lines. The ratio of infectivity of the PBS^{B2} MLV to that of the PBS^{Pro} MLV in a given cell line is a measure of the level of PBS-mediated silencing, after normalizing the ratio for the parental cell line to 1. The control cell lines (control clone 1 and clone 9) had a ratio close to 1, indicating that they exhibited no PBS-mediated restriction. In contrast, clone 5 demonstrated potent PBS-mediated silencing, manifesting as a ratio of ~23-fold (Fig. 4a). Analysis of other clones expressing ZFP809(1–353) showed that all exhibited potent PBS-mediated silencing (Supplementary Fig. 5a). Thus, the expression of ZFP809(1–353) was sufficient to render a differentiated cell resistant to transduction by PBS^{Pro}-using retroviral vectors. To investigate whether ZFP809 silences retroviral expression by binding directly to the integrated provirus, leading to the recruitment of TRIM28, we performed chromatin immunoprecipitations on lysates of clone 5 cells infected with either a restricted (PBS^{Pro}) or an unrestricted variant PBS corresponding to a glutamine tRNA (PBS^{Gln}) MLV (Supplementary Fig. 6). We found that both ZFP809 and TRIM28 are markedly enriched at PBS^{Pro} proviral sites (Supplementary Fig. 6).

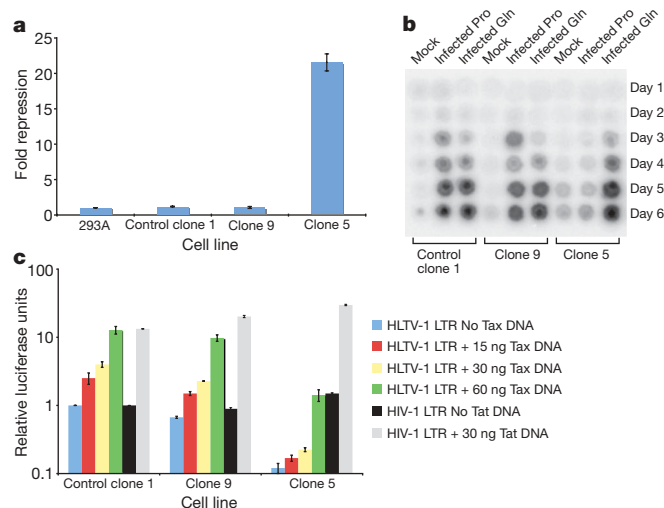


Figure 4 | Expression of ZFP809 in a differentiated cell line causes a potent block to the replication of PBS^{Pro}-using retroviruses. **a**, 293A, control clone 1, clone 9 and clone 5 cell lines were infected with either PBS^{Pro} or PBS^{B2} VSV-G pseudotyped MLV virions expressing the puromycin-resistance gene. Infection efficiency in each cell line was monitored by colony count after puromycin selection. Graph shows the ratio of B2/Pro infection efficiency in each cell line, normalized with 293A = 1. Error bars show \pm s.e.m., $n = 3$. **b**, Control clone 1, clone 9 and clone 5 cell lines were infected at a low multiplicity with replication-competent amphotropic MLV using either PBS^{Pro} (Pro) or PBS^{Gln} (Gln). Viral spread in these cell lines was monitored by analysis of reverse transcriptase activity²² in the culture media every day for 6 days. **c**, HTLV-1 LTR or HIV-1 LTR firefly luciferase reporter activity monitored in control clone 1, clone 5 and clone 9 cell lines, which were simultaneously transfected with increasing amounts of a Tax- or a Tat-expressing vector, as shown. Firefly luciferase values were normalized to *Renilla* luciferase values from a simultaneously transfected *Renilla* control vector. Error bars denote \pm s.e.m., $n = 3$.

To assess whether ZFP809 could also reduce virus replication after authentic retrovirus infection we generated amphotropic MLV virus constructs containing either the restricted PBS^{Pro} or the unrestricted PBS^{Gln}. These viruses were used separately to infect the clone 5 and control clone 1 and clone 9 cell lines at a low multiplicity of infection, and the productive spread of the virus was monitored by measurement of reverse transcriptase activity in the culture media. The amphotropic virus using the wild-type PBS^{Pro} could replicate in the control clone 1 and clone 9 cell lines, but was completely blocked in the clone 5 cell line (Fig. 4b). The amphotropic MLV using PBS^{Gln} was able to spread in all the cell lines.

To investigate whether ZFP809 expression is required for PBS-mediated restriction in embryonic cells, ZFP809 expression was attenuated in F9 ECCs by RNA interference (RNAi). Substantial knockdown of ZFP809 correlated with a complete relief of PBS-mediated restriction (pools 8, 9 and 12) (Supplementary Fig. 7). Messenger RNA analysis also showed lower expression of *Zfp809* in non-restrictive NIH3T3 cells when compared to restrictive ESCs or F9 cells, consistent with the hypothesis that there is a threshold level of ZFP809 required for PBS-mediated restriction (Supplementary Fig. 7).

Having shown that MLV is potently restricted by ZFP809, we wished to determine whether the human pathogen HTLV-1, which also uses a PBS^{Pro}, would also be restricted. A complication to this question is that unlike MLV, HTLV-1 expresses the accessory protein Tax, which recruits co-activators to the long terminal repeat (LTR) and stimulates transcription^{18,19}. It is therefore conceivable that Tax-mediated stimulation of transcription from the LTR might overcome the block induced by ZFP809. To score for transcriptional silencing, we used an HTLV-1 LTR firefly luciferase reporter construct co-transfected with increasing amounts of a Tax-expressing plasmid,

and with a thymidine kinase (TK)-*Renilla* luciferase control plasmid (for normalization) into clone 1, clone 9 and clone 5 cell lines (Fig. 4c). The HIV-1 LTR (which does not contain a PBS^{Pro}) and the coexpression of the Tat transactivator protein were also tested, as a negative control (Fig. 4c). In the case of basal transcription in the absence of Tax, the HTLV-1 LTR reporter construct was silenced 8-fold in clone 5 when compared to control clone 1, and with increasing amounts of Tax this fold silencing was increased, with clone 5 having on average a 14-fold lower LTR activity when compared with control clone 1 (Fig. 4c). These results demonstrate that the HTLV-1 LTR is restricted, and that Tax expression does not abrogate ZFP809-mediated restriction. To demonstrate that ZFP809-mediated restriction requires TRIM28, its levels were stably reduced in clone 5 by RNAi (Supplementary Fig. 8). TRIM28 knockdown resulted in a decrease in PBS binding activity in clone 5 nuclear extracts and a reduction in HTLV-1 LTR restriction in these cells (Supplementary Fig. 8). We note that TRIM28 knockdown in clone 5 cells also led to a decrease in ZFP809 levels, suggesting that ZFP809 protein stability is reliant on formation of the TRIM28–ZFP809 complex (Supplementary Fig. 8).

These findings indicate that ZFP809 is the critical DNA-binding factor in ESCs that recognizes integrated PBS^{Pro}-using retroviral DNAs and silences them through the recruitment of TRIM28. We have shown that viruses using the distinctive Lys1,2 tRNA are also silenced in stem cells in a TRIM28-dependent manner²⁰, and thus a factor must exist that recognizes this distinctive PBS. We speculate that ZFP809 represents the first in a class of PBS-specific retroviral restriction factors that all act through the recruitment of TRIM28 to different PBS elements to silence integrated retroviruses and thereby act as components of a stem-cell-specific intrinsic immunity.

METHODS SUMMARY

EMSA. Double-stranded DNA probes were 5' end-labelled using [γ -³³P]-ATP. Binding reactions were performed using modified Thormell binding buffer⁸ (25 mM HEPES, pH 7.9, 1 mM EDTA, 10% (v/v) glycerol, 5 mM dithiothreitol, 25 ng poly(dI-dC) per μ l, 5 mM NaCl, 5 mM KCl, 3 mM MgCl₂ and 0.1 mM ZnCl₂). Probes (50,000 c.p.m.) were incubated with nuclear extract for 25 min at 30 °C in a total volume of 20 μ l. For supershifts and cold DNA competitions, antibody/cold DNA was added at same time as probe. Binding reactions were analysed by electrophoresis on 4% native polyacrylamide gels⁸.

Ultraviolet crosslinking. Proteins in F9 nuclear extracts were resuspended in modified Thormell binding buffer⁸ and 150,000 c.p.m. ³³P-labelled probe was added to each 50 μ l reaction. Samples were then exposed to 302 nm ultraviolet light for 45 min using a transilluminator (VWR). Competitions were performed with the addition of 100 pmol of unlabelled probe per reaction.

Co-immunoprecipitations. Ten million clone 5 cells were lysed in 750 μ l IPH buffer²¹ (150 mM NaCl, 50 mM Tris-HCl, pH 8, 0.5% Nonidet P-40, 5 mM MgCl₂, 1 mM CaCl₂, 0.5 mM EDTA and protease inhibitors (Complete Protease Inhibitors, Roche)). The lysates were then treated with 100 U DNase (Roche) for 24 h at 4 °C. Ethidium bromide was subsequently added to lysates (50 μ g ml⁻¹) and co-immunoprecipitations were performed using 5 μ g of antibody and a mixture of 20 μ l protein A and 20 μ l protein G Dynabeads (Invitrogen).

Retroviral transduction assays. Virus stocks were collected as previously described⁸. Retroviral preparations were then serially diluted (for titre determination), and added to F9, NIH3T3 or 293A cells (seeded at 3.5×10^3 (F9) and 2×10^3 (NIH3T3 and 293A) cells per cm² the day before transduction) in the presence of 8 μ g of Polybrene per ml. G418 or puromycin-containing selective media was added 24–48 h after infection and colonies were counted 14 days after selection.

LTR reporter assays. 293A cells were seeded at 7,000 cells per well in a 96-well plate and were transfected using Eugene 6 (Roche) with 20 ng HTLV-1 LTR or HIV-1 LTR firefly luciferase reporter plasmid, 5 ng of a TK promoter *Renilla* luciferase control plasmid (phRL-TK, Promega), and a Tax- or Tat-expressing

vector. Luciferase activity was assayed 48 h after infection using dual-luciferase reporter assay (Promega).

Received 17 November 2008; accepted 29 January 2009.

Published online 8 March 2009.

- Teich, N. M., Weiss, R. A., Martin, G. R. & Lowy, D. R. Virus infection of murine teratocarcinoma stem cell lines. *Cell* **12**, 973–982 (1977).
- Barklis, E., Mulligan, R. C. & Jaenisch, R. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* **47**, 391–399 (1986).
- Feuer, G., Taketo, M., Hanecak, R. C. & Fan, H. Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. *J. Virol.* **63**, 2317–2324 (1989).
- Petersen, R., Kempler, G. & Barklis, E. A stem cell-specific silencer in the primer-binding site of a retrovirus. *Mol. Cell. Biol.* **11**, 1214–1221 (1991).
- Yamauchi, M. *et al.* Stem cell factor binding to retrovirus primer binding site silencers. *J. Virol.* **69**, 1142–1149 (1995).
- Wolf, D. & Goff, S. P. Host restriction factors blocking retroviral replication. *Annu. Rev. Genet.* **42**, 143–163 (2008).
- Harada, F., Peters, G. G. & Dahlberg, J. E. The primer tRNA for Moloney murine leukemia virus DNA synthesis. Nucleotide sequence and aminoacylation of tRNA^{Pro}. *J. Biol. Chem.* **254**, 10979–10985 (1979).
- Wolf, D. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**, 46–57 (2007).
- Wolf, D., Cammas, F., Losson, R. & Goff, S. P. Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. *J. Virol.* **82**, 4675–4679 (2008).
- Ayyanathan, K. *et al.* Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev.* **17**, 1855–1869 (2003).
- O'Geen, H. *et al.* Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet.* **3**, e89 (2007).
- Tsuruma, R. *et al.* Physical and functional interactions between STAT3 and KAP1. *Oncogene* **27**, 3054–3059 (2008).
- Koloshia, V. O. & Martin, S. L. *In vitro* properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl Acad. Sci. USA* **94**, 10155–10160 (1997).
- Berg, J. M. & Shi, Y. The galvanization of biology: a growing appreciation for the roles of zinc. *Science* **271**, 1081–1085 (1996).
- Urrutia, R. KRAB-containing zinc-finger repressor proteins. *Genome Biol.* **4**, 231 (2003).
- Friedman, J. R. *et al.* KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* **10**, 2067–2078 (1996).
- Fuks, F. *et al.* The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* **278**, 4035–4040 (2003).
- Felber, B. K. *et al.* The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. *Science* **229**, 675–679 (1985).
- Bex, F. & Gaynor, R. B. Regulation of gene expression by HTLV-I Tax protein. *Methods* **16**, 83–94 (1998).
- Wolf, D., Hug, K. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. *Proc. Natl Acad. Sci. USA* **105**, 12521–12526 (2008).
- Wolf, D. *et al.* Acetylation of β -catenin by CREB-binding protein (CBP). *J. Biol. Chem.* **277**, 25562–25567 (2002).
- Telesnitsky, A., Blain, S. & Goff, S. P. Assays for retroviral reverse transcriptase. *Methods Enzymol.* **262**, 347–362 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Public Health Service grant R37 CA 30488 from the National Cancer Institute. We thank D. Derse, F. Bouamr and E. Barklis for their generosity with reagents. We are grateful to M. de los Santos, H. Nickerson and M. A. Gawinowicz (Protein Core Facility Columbia University) for technical assistance. We would also like to thank P. Adusumilli, M. Banes, M. Bilsky, J. Yamada and co-workers, without whose help this paper would never have been written. D.W. is an Associate, and S.P.G. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions D.W. performed the experiments and analysed the data. D.W. and S.P.G. designed the study and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.P.G. (spg1@columbia.edu).

NEWS

Grants aim to help women

This year, Kristen Williams was able to attend all five days of the annual meeting of the American Physical Society (APS). And she participated in an important tutorial a day earlier — all without fretting about her four-year-old daughter.

Williams, a second-year graduate student in computational materials science at Texas A&M University in College Station, was a beneficiary of Elsevier Foundation's New Scholars grant programme, which is now in its second year. (The foundation, which supports libraries as well as early-career scholars, is an offshoot of Elsevier publishing.) The funds, which she received as an APS grant, allowed her to send her daughter to stay with family in Alabama while Williams spent six days in Pittsburgh, Pennsylvania.

"If it weren't for the grant, I would have skipped the tutorial and gone to the meeting for only a couple days," says Williams. "My daughter would have stayed at home with her father, and I would have been thinking, 'Was it OK, was my husband able to pick her up from school today?' With the grant, I was focused at the meeting and not worrying about her."

The APS's Committee on the Status of Women in Physics is one of five organizations globally that collectively received about \$200,000 recently in grant funds from the foundation. The grants are designed to help finance new programmes that will support women in science. The APS is, in turn, granting



David Ruth: focus on early- to mid-career.

its share of those funds — which it is matching — to individual applicants for child-care costs associated with society conferences. Committee chair Mary Hall-Reno says that the APS has tried other family and child-care programmes for its annual meetings but none has been as effective as this. "It's very flexible," she says. "The child doesn't even have to be at the meeting site."

The other recipients of the New Scholars grants are the

Association for Women in Science, which is developing a coaching and support programme for early- to mid-career women; the Maternal and Childcare Union of Tbilisi, Georgia, which is helping women scientists build networking and other career skills; and the European Molecular Biology Organization, the University of the Pacific and the Museum of New Zealand, all of which received child-care funds for their annual conferences.

David Ruth, foundation executive director, says the grants' focus on childcare, family issues and career advancement grew out of discussions with organizations and agencies internationally, all of which identified early- to mid-career women scientists as most likely to leave the field. They found that concerns about work-life balance were a central reason for the exodus. Ruth says the foundation is trying to spur the development of new programmes to address these conflicts, in hopes of reducing the dropout rate. ■

Karen Kaplan

POSTDOC JOURNAL

The moving blues

I wholeheartedly agree with whomever said that getting married and moving are two of the most stressful things in life. With my husband Brett now in the United States, trying to find his way in his new postdoc position, I am left alone to host the garage sale, sell the car and pack our belongings, while still juggling work and the everyday trials and tribulations of motherhood.

Making a move across the globe is both an exciting and a terrifying prospect at any time, but doing it with a three-year-old brings many

other considerations. While Brett struggles to find us an affordable house in a child-friendly neighbourhood and organize suitable day care in the United States, here I try to calm an upset toddler who is missing his daddy, has lost 90% of his toys in the past week and is now sleeping on a mattress.

And sizeable moving costs, which we have to cover, have exacerbated our already shaky financial circumstances. I know we're not alone. This financial crunch is common for postdocs taking a position overseas, and we can only

have faith that the move will reap financial and professional rewards in the future.

Add to the mix the inevitable sadness of leaving behind good friends and familiar surroundings (and, of course, the tropical Australian climate!), and I think I am justified in feeling a little melancholy. And yet I remain excited. I eagerly anticipate our new adventures — even the possibility of celebrating a snowy Christmas. ■

Joanne Isaac was a postdoc in climate-change effects on biodiversity at James Cook University, Townsville, Australia.



IN BRIEF

Grant funding cuts

Researchers may have trouble securing funding from US foundations this year, according to a survey from the New York-based non-profit Foundation Center. Nearly two-thirds of the 1,200 foundations that responded plan to cut the number and/or size of grants. Less-established organizations are at a particular disadvantage, says Steven Lawrence, the centre's senior director of research. However, 14% of the foundations plan to make 'exceptional' grants as a result of the downturn, funding more humanitarian programmes or initiatives outside their usual domain.

Solar R&D in Arizona

Science Foundation Arizona, a non-profit organization that helps to diversify the state's economy, is creating the Solar Technology Institute (STI) with a \$4 million investment. The new partnership of universities and energy industries will launch five solar R&D projects, says STI co-director Dick Powell. One of the project leaders, Roger Angel, a University of Arizona astronomer and optical scientist, will hire about six researchers to produce commercially viable satellite-based reflectors to concentrate sunlight onto photovoltaic cells. Other projects will focus on the use of nanotechnology in energy storage, smart-grid management and reduction of photovoltaic-cell cost.

University research

The quality and amount of a university's research is linked to three factors, a new study reports. These are research funding; ability to decide on its own programmes, hires and budgets; and the level of competition it faces for resources, faculty and students. *The Governance and Performance of Research Universities: Evidence from Europe and the US*, a draft by the non-profit National Bureau of Economic Research, states that if one of the three is missing, research output and quality drops. More research spending spurs more patents if universities are autonomous and have to compete for grants, faculty and students — but whether the relationship is causal or correlative is unclear. Proximity to regions of high innovation and productivity also raises the quality and amount of a university's research output, the report finds.



State of energy

New Mexico, with its national labs and natural resources, is poised to become a central player in the US race for energy independence. **Paul Smaglik** reports.

The state of New Mexico already ranks top in the United States for PhDs per capita, with roughly one for every 300 of its 1.8 million citizens. And if its institutions and researchers have their way, New Mexico will become an even stronger scientist magnet in the coming years as energy research, the state's speciality, continues to be a major focus of state and federal government agendas.

New Mexico hosts two national Department of Energy laboratories, Sandia and Los Alamos, which collectively boast a \$4.4 billion budget, 18,000 employees and 3,700 PhDs. The University of New Mexico, New Mexico State University and the New Mexico Institute of Mining and Technology (known as New Mexico Tech) also devote substantial resources to energy research. Private partners with both the universities and the national labs — oil and car companies, for example — extend the research capacity even further.

This potential, matched with the new US administration's recently unveiled research and development agenda, presents the state with a prime opportunity. President Barack Obama's economic stimulus package includes a whopping \$32.7 billion for the Department of Energy to distribute in grants. That includes \$4 billion for energy-saving 'smart grid' programmes — money that New Mexico is

eagerly competing for, according to Thomas Bowles, an astrophysicist and former chief science officer at the Los Alamos National Laboratory who is currently working as science adviser to New Mexico governor Bill Richardson. As part of his efforts, Richardson hopes to secure additional funds for a proposed 'Green Grid Initiative' to update energy creation, storage and transmission throughout New Mexico, a project that could bring more energy research jobs to the state.

The first phase of the Green Grid Initiative will enhance the state's bid for at least one of Obama's energy research centres of excellence, paid for by smart-grid funds. Private enterprise is also part of the equation.

"We're looking to attract business partners into the state for clean manufacture of energy, then refine and improve these processes," Bowles says. He adds that within five years, New Mexico wants to demonstrate to venture capitalists the capacity it has for producing, storing and transporting clean energy.

New Mexico has the requisite natural resources, with the second highest potential for solar production in the United States, behind only Nevada, and the twelfth highest for wind. The state's universities and national laboratories are already working on ways to reduce energy use, store unused power generated at off-peak

times and move that power more efficiently around a grid.

Sandia has worked with the US Air Force to develop a better solar cell to power satellites. Sandia scientists may also evaluate the nation's largest solar farm at Nellis Air Force Base in Nevada and study it to build similar facilities in New Mexico.

But a less obvious potential energy source lies in the sea — or, at least, the sea that used to be. Millions of years ago, New Mexico was covered by a salt-water ocean. Brackish pools — the perfect home for algae — persisted as the sea retreated. Abhaya Datye, professor of chemical and nuclear engineering at the University of New Mexico in Albuquerque, has a five-year, \$2.5-million grant from the National Science Foundation to help turn algae into biofuels by improving the catalytic process that turns organic matter into biofuels, triggered by the Sun.

Datye is drawing on research interactions with Sandia, major automakers and fossil-fuel companies, with input from students and postdocs. He says that students and postdocs from Germany and Denmark as well as from Iowa State University, the University of Wisconsin, Madison, and the University of Virginia will rotate among labs involved in the grant. They will also work with industrial partners, including Haldor Topsoe, a large Danish catalyst company.

Light reduction

Alternative energy sources are not the only focus; New Mexico scientists are looking to reduce energy consumption as well. Stephen Hersee, a professor of materials at the University of New Mexico, has a head start. He



"Resilience and adaptability of systems tend to be overlooked." —

Geoffrey West



Light industry: (from left) Sandia Science and Technology Park, DOE Center for Integrated Nanotechnologies, Sandia MicroLab and Nellis Air Force Base.

SANDIA NATIONAL LABORATORIES

SANDIA NATIONAL LABORATORIES

is part of a collaboration that, last September, received a grant from the National Science Foundation that could total \$30 million over 10 years. The collaboration's goal is to reduce the amount of energy used for lighting, beyond the capability of compact fluorescent light bulbs. The next step is improving the efficiency of light-emitting diodes. Hersee estimates that changing all US light bulbs to solid state would reduce the country's energy bill by 20%. And as solid-state bulbs don't contain mercury, they don't present the same disposal problems caused by compact fluorescents. Hersee and colleagues' group is large and diverse, with some 20 researchers across three centres plus another 80–100 undergraduate and graduate researchers. Sandia is also involved and the collaboration is seeking 12 industrial partners.

In a prime example of a government–private collaboration, Sandia and Goodyear Tires of Akron, Ohio, are working to make tyres more fuel efficient. They are searching for greener ways to cure rubber, a costly and energy-draining part of the manufacturing process. “People ask me all the time, ‘What do tyres have to do with nuclear weapons?’” says Hal Morgan, a Sandia scientist who heads the Goodyear collaboration. Both involve engineering expertise and computational modelling, he says — capacities that may help companies build more energy-efficient products. About 20 Sandia scientists work with Goodyear, which provides \$5 million a year for Sandia's research: some \$35 million since the collaboration began 15 years ago.

Fuel tank

Even if the world can increase energy production through biofuels, solar power, wind and hydropower, a major hurdle remains — storage. That's why the Los Alamos National Laboratory is spending \$250 million on basic science research to improve energy storage and transmission. “It's easy to double generation,” says

Terry Wallace, the Los Alamos National Laboratory's associate director of science, technology and engineering. “It's hard to put it on the grid. Where do you put storage?”

Practical answers may appear in a ‘green’ development named Mesa del Sol outside Albuquerque. Developer Forest City Residential Group wants to improve solar-power generation technology and find ways to store excess power, then return it to the grid. Sandia and partners are looking into some options. Concentrated solar power might provide a way to store thermal energy using molten salt, and compressed-air technology might allow excess wind power to be driven into the ground during off-peak hours, then released to drive turbines later when the wind subsides.

One short-term answer for improving solar technology in Mesa del Sol may lie in Hawaii, where fossil fuels are even more expensive than in the continental United States. The island of Lanai is launching a new kind of solar cell this summer, with battery storage for excess energy produced. Sandia scientists such as Juan Torres will be monitoring and evaluating that project to see how it scales up.

Taking technology from a relatively small test site, such as Lanai or the 38,000 houses planned for Mesa del Sol, to larger urban areas and nationwide networks poses perhaps the biggest challenge, say scientists engaged at all levels from production to storage. That's where the theoretical physicists come in. The Los Alamos National Laboratory and Sandia have the computing power necessary to model the complex interaction of energy production, storage and distribution. And the Department of Homeland Security is using

Playas, a defunct mining town owned by New Mexico Tech, to do real-world modelling of how a terrorist threat to one node of the energy system affects the entire grid.

Elements of chaos

“When we look at these types of systems, one thing that tends to be overlooked

from a long-term view is the resilience, robustness and adaptability of these systems,” says Geoffrey West, president of the Santa Fe Institute in New Mexico.

The trick is anticipating the possible elements of chaos — small perturbations in one part of the system that can have major consequences on other parts. One real-world example is the evidence of Internet ‘spies’ lurking on some nodes of the US energy grid. Those spies could quickly turn into energy terrorists, disabling the power grid or turning a ‘smart grid’ into an information roadblock. Such calamities — both

anticipated and unforeseen — are why the Santa Fe Institute collaborates with various stakeholders in projects such as energy. “The very nature of the work we do here, even though it comes under the rubric of basic research, brings us face to face with all these major problems society faces,” West says.

New Mexico, with its multiple players in the research network — national labs, state universities, research institutes and private companies — is well placed to create a new energy network. Its challenges will be to develop each of those nodes, connect them and scale them up: work that should provide ample research opportunities for the foreseeable future.

Paul Smaglik is a freelance writer based in Milwaukee, Wisconsin.



Stephen Hersee: Changing all light bulbs to solid state would reduce America's energy bill by 20%.

US AIR FORCE

En passant

Family ties.

**Michalis Barkoulas
and Gemma Bilsborough**

"White or black?" asked Franck with a quivering voice while he was poured another glass of French brandy.

"I'd rather play black," little Frank replied.

Franck had expected that answer, so he wasn't surprised. "Sure. Why don't you assemble the pieces and I can take care of the music," he said, and started walking towards the stereo. He looked confused, as if he didn't know what melodies would best suit the occasion. He was quite nervous too. This was the moment he had been waiting for for a long time, and, despite all the forethought, he still felt like a fish out of water. "What about some funky jazz?" Franck asked.

"Why not? That's my favourite," little Frank replied, and they both laughed with gusto.

"I think it's reasonable to expect a tie," Franck pointed out in an attempt to further break the ice, and looked at little Frank with a cunning smile. Then he tapped the edge of the table, 'pam-pam', officially denoting the beginning of the game. As white plays first, Franck decided to kick off with a standard move: pawn to E3. Little Frank wasted no time and confidently moved a pawn to B6.

Franck was a proficient chess player. He had an everlasting, almost erotic relationship with his expensive, silver board. He also had a passion for reading books about the game: books about chess tactics and chess problems. Many books. Even on the psychology of the game. He was a psychologist by trade. Like little Frank, he would always choose black. Not because game theory suggests that black is advantageous, but because Franck was waiting to perform psychoanalysis. "I want you exposed first," he used to say. "After all, that's my job!" He strongly believed that the way you play reveals a lot about your personality. If you follow an offensive tactic, you are aggressive and a risk-taker. If you take your time over each move, then you are careful and diffident. After so many years of correspondence, he could finally spend a few precious minutes together with little Frank. And playing chess was the perfect chance for Franck to remain calm and get an idea of what kind of a person little Frank was.

"Bishop to C4," Franck announced, and then immediately lit his pipe. "I forgot to ask you, do you smoke, Frank?"

Frank moved his knight to C6. "Yes, I always have, but only in secret. They don't want us to die young like Dolly. Dollies are bad publicity for the company, they say."

The arrangement between Franck and the cloning company was simple and not negotiable. They request your genes to develop their protocols. You need the money, so you sign the ED (ethical dilemma) form.

If the cloning



works, you receive an automatic confirmation phone call. You do get to choose the name, but, apart from that, no further action is required. In fact, the company stipulates that you don't make contact with your clone, so as to avoid emotional repercussions. Franck still remembered the day he found out about the existence of little Frank.

"Please spell out the name of your clone after the beep. May I remind you that, according to the terms of our legal agreement, it is compulsory to choose a name for your clone that is distinct from your own name. Beeeep." But Franck was an old-school romantic and found that notion ridiculous. "How can I choose a different name for ... myself?" "FRANK," Franck said, "F-R-A-N-K."

Franck moved his queen to F3. "Look out, little Frank, you are in trouble there!"

But Frank didn't seem to bother and quickly moved his bishop to B7. "What a careless move," Franck thought. "He looks aggressive, but what do you expect from a child who has been on his own from the very early days?" But soon, another thought interrupted his logic: "Gosh, he looks so much like me when I was younger."

There were only two things Franck hated in other people: stinginess and lies. Not disliked, but hated, venomously. "Stingy people are shallow," he used to say, and he wouldn't talk to a person again if he realized they were stingy. But, on this occasion, he was delighted to see that little Frank was not penurious. "A full bottle of vintage Napoleon brandy when he is still a psychology student! What a generous gift this is." He had no doubts about Frank telling

lies either. "Loathing lies is so strong in my mind that I am sure it must be encoded in my genes," Franck thought proudly. So, little Frank must surely be honest. "His name is Frank, after all," he observed.

"Queen to F7 and ... I am afraid it's over!" Franck said apologetically, declaring a precocious finale to the game.

"Oh no! How did I not see that coming?" little Frank replied. "I think I underestimated your Napoleon opening," he added, and sipped down the remaining brandy in his glass.

"Will I see you again ... Dad?" he asked hurriedly, but they both realized how funny that statement was, and avoided making promises.

It was on the runaway train when Frank felt remorseful over stealing the brandy from an off-licence and the black king from Franck's favourite board. He wept with frustration. His life flashed in front of his eyes: several unsuccessful attempts to get into university, no real family, many fears and lies, and no hopes. Fuelled with desperation, he slowly opened his fist, releasing the king from his grasp. "Check-mate," he whispered, and chewed the conium leaves. ■

Michalis Barkoulas and Gemma Bilsborough are based at the Department of Plant Sciences at the University of Oxford, where they investigate leaf development in crucifer plants. They dreamt that they devised this story during a short chess break in the lab, waiting for a long PCR reaction to finish.

JACEY